# This Page Is Inserted by IFW Operations and is not a part of the Official Record

### **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.

### (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

## (19) World Intellectual Property Organization International Bureau



### 

### (43) International Publication Date 14 December 2000 (14.12.2000)

### **PCT**

## (10) International Publication Number WO 00/75279 A2

(51) International Patent Classification7:

C12N

(21) International Application Number: PCT/US00/15728

(22) International Filing Date: 6 June 2000 (06.06.2000)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data: 09/328,925

9 June 1999 (09.06.1999) US

(71) Applicant (for all designated States except US): RE-GENTS OF THE UNIVERSITY OF MICHIGAN [US/US]; 3003 South State Street, Wolverine Tower, Ann Arbor, MI 48109-1280 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KURACHI, Kotoku [JP/JP]; 2211 Avalon Place, Ann Arbor, MI 48104 (US). KURACHI, Sumiko [JP/JP]; 2211 Avalon Place, Ann Arbor, MI 48104 (US).

- (74) Agents: CARROLL, Peter, G. et al.; Medlen & Carroll, LLP, Suite 2200, 220 Montgomery Street, San Francisco, CA 94104 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

#### Published:

 Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A2

(54) Title: NUCLEOTIDE SEQUENCES FOR GENE REGULATION AND METHODS OF USE THEREOF

(57) Abstract: The invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest. In particular, the invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner and/or in a liver-specific manner. The invention further provides methods of using the regulatory nucleic acid sequences provided herein for age-related and/or liver-specific expression of nucleotides sequences of interest. The invention also provides host cells and transgenic non-human animals which harbor the regulatory nucleic acid sequences of the invention. The compositions and methods of the invention are useful in regulating expression of a nucleotide sequence of interest in an age-related and/or liver-specific manner.

## NUCLEOTIDE SEQUENCES FOR GENE REGULATION AND METHODS OF USE THEREOF

This work was made, in part, with Government support by the National Institutes of Health grant numbers HL38644 and HL53713. The Government has certain rights in this invention.

### FIELD OF THE INVENTION

5

10

15

20

25

30

The invention relates to nucleic acid sequences which regulate expression of a nucleotide sequence of interest. In particular, the invention relates to nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner and/or in a liver-specific manner. The invention further relates to methods of using the regulatory nucleic acid sequences provided herein for age-related and/or liver-specific expression of nucleotides sequences of interest. The invention also relates to host cells and to transgenic non-human animals which harbor the regulatory nucleic acid sequences of the invention. The compositions and methods of the invention are useful in regulating expression of a nucleotide sequence of interest in an age-related and/or liver-specific manner.

### **BACKGROUND OF THE INVENTION**

A multitude of human diseases (e.g., thrombosis, cardiovascular diseases, diabetes, Alzheimer's disease, cancer, osteoporosis, osteoarthritis, Parkinson's disease, dementia) are associated with increasing age and result in serious effects on the quality of life and on the life expectancy of individuals suffering from such diseases. Other diseases (e.g., cirrhosis, primary and metastatic neoplasia, Wilson disease, hepachromatosis, infectious hepatitis, hepatic necrosis, Gilbert disease, Criggler-Najar disease) which afflict the liver also have serious clinical manifestations and are responsible for high morbidity and mortality.

The treatment of age-related diseases (i.e., diseases whose prevalence and/or severity of clinical manifestations increases with the age of the patient) and diseases afflicting the liver focuses on the alleviation of the general symptoms of the disease using one or a combination of two modalities, i.e., non-pharmacological treatment and pharmacological treatment. Non-pharmacological treatment include, for example, periods of bed rest and dietary changes. Non-pharmacological treatment is often used as an adjunct to

pharmacological treatment which involves the use of drugs. Unfortunately, many of the commonly used pharmacological agents have numerous side effects and their use is further exacerbated by the non-responsiveness by many patients with severe disease, who, paradoxically, are in most need of treatment. Both non-pharmacological and pharmacological treatments provide unsatisfactory approaches to treating age-related and liver-associated diseases because these approaches are often ineffective, their effects are inconsistent, and are directed to alleviating the general symptoms of disease, rather than to specifically addressing the source of morbidity and mortality. Moreover, no suitable animal models are currently available to rationally design drugs which target specific biochemical and physiological pathways which are associated with age-related and with liver-associated diseases.

What is needed are methods for age-related and liver-specific gene expression and models for age-related and liver-specific diseases.

### SUMMARY OF THE INVENTION

5

10

15

20

25

30

The invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner, as well as nucleic acid sequences which direct liver-specific expression of a gene of interest. Further provided by the invention are transgenic animals which may be used as models for age-related and/or liver specific diseases.

In one embodiment, the invention provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without intending to limit the invention to any particular type or source of nucleic acids sequence of interest, in a preferred embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, fibrinolytic pathway factors and inhibitors, PEA-3 protein, PEA-3 related proteins including Ets family transcriptional factors, β-galactosidase, and luciferase. While it is not intended that the invention be restricted to any particular type or source of promoter sequence, in an alternative preferred embodiment, the promoter sequence is selected from human factor IX

promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>7</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. It is not contemplated that the invention be limited to any particular age regulatory sequence which is a portion of SEQ ID NO:1. However, in another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:31 is selected from SEQ ID NO:2, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, and SEQ ID NO:38. Without intending to limit the invention to any particular rage regulatory sequence which s a portion of SEQ ID NO:3, in yet another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:3 is selected from SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, and SEQ ID NO:61.

5

10 .

15

20

25

30

Also provided by the invention is a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without intending to limit the invention to the environment in which the host cell is contained, in one preferred embodiment, the host cell is comprised in a tissue or organ in a living animal. In another preferred embodiment, the host cell is a gamete. In yet another preferred embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

The invention also provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without limiting the invention to the type or source of the nucleic acid sequence of interest, in one preferred embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, fibrinolytic pathway factors and inhibitors, PEA-3 protein, PEA-3 related proteins including Ets family transcriptional factors, β-galactosidase, and luciferase. While it is not intended that the invention be limited to the type or source of the promoter sequence, in an alternative

preferred embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>7</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. Though it is not contemplated that the invention be limited to the portion of SEQ ID NO:1 which has age-related regulatory activity, in another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:1 is selected from SEQ ID NO:2, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, and SEQ ID NO:38. Without intending to limit the invention the portion of SEQ ID NO:3 which has age-related regulatory activity, in yet another preferred embodiment, the age regulatory sequence which is a portion of SEQ ID NO:3 is selected from SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58; SEQ ID NO:59, SEQ ID NO:60, and SEQ ID NO:61.

5

10

15

20

25

30

Also provided herein is a host cell containing recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3. Without intending to limit the invention to the environment in which the host cell is contained, in one preferred embodiment, the host cell is comprised in a tissue or organ in a living animal. In an alternative preferred embodiment, the host cell is a gamete. In another preferred embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

The invention also provides a method, comprising: a) providing: i) a cell, ii) a nucleic acid sequence of interest, iii) a promoter sequence, and iv) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the one or more age regulatory sequences to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell. Without intending to limit the treated cell to any particular environment, in one preferred embodiment, the treated cell is comprised in a tissue or organ in a living animal.

The invention further provides a substantially purified nucleic acid sequence comprising a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof.

Also provided herein is a substantially purified nucleic acid sequence comprising a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof.

5

10

15

20

25

30

Also provided by the present invention is a substantially purified nucleic acid sequence comprising a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof. In one embodiment, the portion is SEQ ID NO:2. In an alternative embodiment, the portion is selected from SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37 and SEQ ID NO:38.

The invention also provides a substantially purified nucleic acid sequence comprising a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof. In one embodiment, the portion is selected from SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, and SEQ ID NO:61.

Also provided herein is a substantially purified nucleic acid sequence which hybridizes under stringent hybridization conditions with SEQ ID NO:1 or with the complement thereof, wherein the nucleic acid sequence is characterized by having age-related regulatory activity, and by having greater than 63% and less than 100% homology to the SEQ ID NO:1.

The invention also provides a substantially purified nucleic acid sequence which hybridizes under stringent hybridization conditions with SEQ ID NO:3 or with the complement thereof, wherein the nucleic acid sequence is characterized by having age-related regulatory activity, and by having greater than 60% and less than 100% homology to the SEQ ID NO:3.

The invention additionally provides a recombinant expression vector comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof.

Also provided herein is a recombinant expression vector comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof.

The invention also provides a transgenic cell comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:1 and of the complement thereof. In one embodiment, the nucleotide sequence is operably linked to a promoter and to a nucleic acid sequence of interest. In a preferred embodiment, the transgenic cell is comprised in an animal. In a more preferred embodiment, the nucleic acid sequence of interest is expressed in an age-related manner in the transgenic cell.

5

10

15

20

25

30

The invention additionally provides a transgenic cell comprising at least a portion of a nucleotide sequence selected from a functional homolog of SEQ ID NO:3 and of the complement thereof. In one embodiment, the nucleotide sequence is operably linked to a promoter and to a nucleic acid sequence of interest. In a preferred embodiment, the transgenic cell is comprised in an animal. In a more preferred embodiment, the nucleic acid sequence of interest is expressed in an age-related manner in the transgenic cell.

The invention also provides a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; iv) SEQ ID NO:1; and v) SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, the SEQ ID NO:1 and the SEO ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell. In one embodiment, the cell expresses a recombinant protein identified as SEQ ID NO:47. In an alternative embodiment, the cell is selected from HepG2 cell, fibroblast cell, myoblast cell, and endothelial cell. In another embodiment, the cell is a fertilized egg cell, and the transgenic cell is a transgenic fertilized egg cell. In a preferred embodiment, the method further comprises d) introducing the transgenic fertilized egg cell into a non-human animal and permitting the animal to deliver progeny containing the transgene. In a more preferred embodiment, the progeny is characterized by age-related expression of the nucleic acid sequence of interest. In an alternative more preferred embodiment, the progeny is characterized by liver-specific expression of the nucleic acid sequence of interest. In another preferred embodiment, the fertilized egg cell is derived from a mammal of the order Rodentia. In a more preferred embodiment, the fertilized egg cell is a mouse fertilized egg cell. In yet another embodiment, the promoter is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK

promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>7</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In a further embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, fibrinolytic pathway factors and inhibitors, PEA-3 protein, PEA-3 related proteins including Ets family transcriptional factors, β-galactosidase, and luciferase.

5

10

15

20

25

30

The invention also provides a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; iv) a portion of SEQ ID NO:1; and v) a portion of SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, the portion of SEQ ID NO:1 and the portion of SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

Additionally provided by the invention is a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) SEQ ID NO:1; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the SEQ ID NO:1 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

Also provided herein is a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a portion of SEQ ID NO:1; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the portion of SEQ ID NO:1 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

The invention further provides a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

Further provided by the invention is a method for expressing a nucleic acid sequence of interest in a cell, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a portion of SEQ ID NO:3; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the portion of SEQ ID NO:3 to produce a transgene; and c) introducing the transgene into the cell to produce a transgenic cell under conditions such that the nucleic acid sequence of interest is expressed in the transgenic cell.

5

10

15

20

25

30

The invention further also provides additional sequences derived from the hFIX gene. In particular, the invention provides a substantially purified nucleic acid sequence comprising at least a portion of SEQ ID NO:93. In one embodiment, the portion has age-related regulatory activity. In an alternative embodiment, the portion is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144. In a preferred embodiment, the portion is SEQ ID NO:91.

Also provided herein is a substantially purified nucleic acid sequence which hybridizes under high stringency hybridization conditions with a nucleotide sequence selected from SEQ ID NO:91, the complement of SEQ ID NO:91, SEQ ID NO:93, and the complement of SEQ ID NO:93. In one embodiment, the nucleic acid sequence has agerelated regulatory activity.

The invention also provides a substantially purified nucleic acid sequence comprising a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof.

The invention further provides expression vectors containing hFIX sequences. In one embodiment, the invention provides a recombinant expression vector comprising in operable

5

10

15

20

25

30

combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) an agerelated regulatory sequence selected from SEQ ID NO:93 and portions thereof. In a preferred embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In an alternative embodimnt, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRa promoter, P450 family promoters, GAL7 promoter, T<sub>2</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In yet another alternative embodiment, the portion of SEO ID NO:93 is selected from SEO ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144. In a more preferred embodiment, the portion is SEQ ID NO:91. In another embodiment, the expression vector further comprises in operable combination an age-related regulatory sequence selected from SEQ ID NO:1 and portions thereof.

The invention further provides cells containing hFIX-derived sequences. In particular, the invention provides a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In an alternative embodiment, the host cell is a gamete. In another alternative embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

5

10

15

20

25

30

The invention additionally provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of an age-related regulatory sequence selected from SEO ID NO:93 and portions thereof. In one embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human \(\alpha\)1-antitrypsin. antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In an alternative embodiment, the promoter sequence is selected from human factor IX promoter. cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRa promoter, P450 family promoters, GAL7 promoter, T<sub>2</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In yet another embodiment, the portion of SEQ ID NO:93 is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEO ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144. In yet a further embodiment, the portion is SEQ ID NO:91. In another embodiment, the expression vector further comprises in operable combination an age-related regulatory sequence selected from SEQ ID NO:1 and portions thereof.

Also provided herein is host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In another embodiment, the host cell is a gamete. In yet

another embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

Also provided herein are methods for using hFIX-derived sequences. For example, the invention provides a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the age-related regulatory sequence to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell. In one embodiment, the treated cell is comprised in a tissue or organ in a living animal.

5

10

15

20

25

30

The invention also provides a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a functional homolog of an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the functional homolog to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell.

In addition to sequences from the hFIX gene, the invention provides sequences derived from the hPC gene. In particular, the invention provides substantially purified nucleic acid sequence comprising a nucleotide sequence selected from at least a portion of SEQ ID NO:85, and at least a portion of SEQ ID NO:92. In one embodiment, the portion has activity selected from age-related regulatory activity and regulatory activity. In an alternative embodiment, the portion is selected from SEQ ID NO:88, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162; SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:87, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:179, SEQ ID NO:171, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:181, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEO ID NO:184, SEO ID

NO:185, SEQ ID NO:186, SEQ ID NO:187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:89, and SEQ ID NO:90. In another alternative embodiment, the portion is selected from SEQ ID NO:89 and SEQ ID NO:90.

5

10

15

20

25

30

Also provided herein is a substantially purified nucleic acid sequence which hybridizes under stringent hybridization conditions with a nucleotide sequence selected from SEQ ID NO:92, the complement of SEQ ID NO:95, the complement of SEQ ID NO:85, SEQ ID NO:89, the complement of SEQ ID NO:89, SEQ ID NO:90, and the complement of SEQ ID NO:90. In one embodiment, the nucleic acid sequence has activity selected from age-related regulatory activity and regulatory activity.

The invention additionally provides a substantially purified nucleic acid sequence comprising a functional homolog of a sequence having activity selected from age-related regulatory activity and regulatory activity, the sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90.

Also provided by the invention is a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the expression vector of Claim 38, wherein the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In another embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>7</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In yet another embodiment, the portion is

selected from SEQ ID NO:88, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:87, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:197, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:197, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:89, and SEQ ID NO:90. In a further embodiment, the portion is selected from SEQ ID NO:89 and SEQ ID NO:90.

The invention also provides a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In another embodiment, the host cell is a gamete. In a further embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

25

30

5

10

15

20

The invention also provides a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin,

5

10

15

20

25

30

protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase. In another embodiment, the promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRa promoter, P450 family promoters, GAL7 promoter, T<sub>2</sub> promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter. In a further embodiment, the portion is selected from SEQ ID NO:88, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:87, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:89, and SEQ ID NO:90. In an alternative embodiment, the portion is selected from SEQ ID NO:89 and SEQ ID NO:90.

Further provided herein is a host cell containing a recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) a functional homolog of a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90. In one embodiment, the host cell is comprised in a tissue or organ in a living animal. In an alternative embodiment, the host cell is a gamete. In yet another embodiment, the host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.

Also provided herein are methods for using hPC sequences. In particular, the invention discloses a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the nucleotide sequence to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell. In one embodiment, the treated cell is comprised in a tissue or organ in a living animal.

Also provided herein is a method for expressing a nucleic acid sequence of interest, comprising: a) providing: i) a cell; ii) a nucleic acid sequence of interest; iii) a promoter sequence; and iv) a functional homolog of a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, the nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90; b) operably linking the nucleic acid sequence of interest, the promoter sequence, and the functional homolog to produce a transgene; and c) introducing the transgene into the cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in the treated cell.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

5

10

15

20

25

30

Figure 1 shows the structure of eleven exemplary human FIX minigene expression constructs and relative *in vitro* transient expression activities (ng hFIX/10<sup>6</sup> cells/48 hr).

Figure 2 shows graphs of longitudinal analyses of transgenic mice which carry -416FIXm1 (A), -416FIXm1/1.4 (B), -590FIXm1 (C), -679FIXm1 (D), and -770FIXm1 (E) expression vectors and which produce high initial prepubertal, but rapidly decreasing, hFIX expression levels with age.

Figure 3 shows a Northern blot of human FIX mRNA levels (A) and a gel showing hFIX transgene DNA levels as determined by multiplex PCR analysis (B) in the livers and tails of animals carrying -416FIXm1.

Figure 4 shows graphs of longitudinal analysis of transgenic mice which carry -802FIXm1 (A), -802FIXm1/1.4 (B), -2231FIXm1 (C), -2231FIXm1/1.4 (D) and -416FIXm1/AE5' (E) expression vectors and which produce hFIX at stable and increasing levels with age.

Figure 5 shows a Northern blot of transgenic mice carrying -802FIXm1 and -802FIXm1/1.4 expression vectors.

5

10

15

20

25

30

Figure 6 is a gel of a gel electrophoretic mobility shift assay using mouse liver nuclear extract (NEs) from three different age groups, and using double-stranded oligonucleotides containing a PEA-3 element nucleotide sequence spanning from nt -797 to -776 of the hFIX gene (A), and using a competition assay for <sup>32</sup>P-labelled double stranded oligonucleotides containing the PEA-3 nucleotide sequence (B).

Figure 7 is a Northern blot showing tissue specificity of hFIX expression in transgenic mice carrying -416FIXm1 (A) and -802 FIXm1 (B) expression vectors.

Figure 8A-E shows the nucleotide sequence (SEQ ID NO:4) of, and eight amino acid sequences (SEQ ID NOs:5 to 12) which together form, the human factor IX (GenBank accession number K02402). The initiation transcription site (nucleotide 1) and the poly-A addition site (nucleotide 32,757) are identified by solid circles. The solid vertical arrows indicate the intron-exon splice junction. The five Alu repetitive sequences have been underlined, while the 5-base insert in intron A and the AATAAA sequence in exon VIII are boxed. The cleavage or termination site at the 3' end of the gene (CATTG) is underlined with a dashed line.

Figure 9 shows the cDNA sequence (SEQ ID NO:13) (A) and encoded polypeptide sequence (SEQ ID NO:47) (B) of mouse PEA-3 (GenBank accession number X63190).

Figure 10 A-D shows the cDNA sequence (SEQ ID NO:42) of the human α1-antitrypsin gene (GenBank accession number K02212).

Figure 11 shows the DNA sequence (SEQ ID NO:43) of human antithrombin III (GenBank accession number A06100).

Figure 12 shows the cDNA sequence (A) (SEQ ID NO:49) (GenBank accession number X02750) and genomic DNA sequence (B) (SEQ ID NO:50) (GenBank accession number M11228) of human protein C.

Figure 13 (A-E) shows the nucleic acid sequences (SEQ ID NOs:76-83) of exemplary homologs of AE3' (SEQ ID NO:3).

Figure 14 shows the nucleotide sequence (from nt -1462 to nt +1; SEQ ID NO:85) which is located at the 5'-end of the human protein C gene.

Figure 15 shows the structure of eight exemplary human protein C minigene expression constructs.

Figure 16 shows the relative *in vitro* transient expression activities for five exemplary human protein C minigene expression constructs.

Figure 17 shows graphs of longitudinal analyses of transgenic mice which carry -1462hPCm1 (A), -82hPCm1 (B), and AE5'/-1462hPCm1/AE3' (C) expression vectors.

### **DEFINITIONS**

5

10

15

20

25

30

To facilitate understanding of the invention, a number of terms are defined below.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated nucleic acid sequence" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural state, or when obtained from its actual source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, an isolated nucleic acid sequence comprising SEQ ID NO:1 includes, by way of example, such nucleic acid sequences in cells which ordinarily contain SEQ ID NO:1 where the nucleic acid sequence is in a chromosomal or extrachromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid sequence may be present in single-stranded or double-stranded form. When an isolated nucleic acid sequence is to be utilized to express a protein, the nucleic acid sequence will contain (at a minimum) at least a portion of the sense or coding strand (i.e., the nucleic acid sequence may be single-stranded). Alternatively, it may contain both the sense and antisense strands (i.e., the nucleic acid sequence may be double-stranded).

As used herein, the term "purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated. An

"isolated nucleic acid sequence" is therefore a purified nucleic acid sequence. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated.

The term "recombinant" when made in reference to a DNA sequence refers to a DNA sequence which is comprised of segments of DNA joined together by means of molecular biological techniques. The term "recombinant" when made in reference to a polypeptide sequence refers to a polypeptide sequence which is expressed using a recombinant DNA sequence.

5

10

15

20

25

30

As used herein, the terms "vector" and "vehicle" are used interchangeably in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another.

The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid sequences necessary for expression in prokaryotes include a promoter, optionally an operator sequence, a ribosome binding site and possibly other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

The term "transgenic" when used in reference to a cell refers to a cell which contains a transgene, or whose genome has been altered by the introduction of a transgene. The term "transgenic" when used in reference to a tissue or animal refers to a tissue or animal, respectively, which comprises one or more cells that contain a transgene, or whose genome has been altered by the introduction of a transgene. Transgenic cells, tissues and animals may be produced by several methods including the introduction of a "transgene" comprising nucleic acid (usually DNA) into a target cell or integration of the transgene into a chromosome of a target cell by way of human intervention, such as by the methods described herein.

A "non-human animal" refers to any animal which is not a human and includes vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc. Preferred non-human animals are selected from the order Rodentia. The term "order Rodentia" refers to rodents *i.e.*, placental mammals (class Euthria) which include the family Muridae (e.g., rats and mice), most preferably mice.

The term "nucleotide sequence of interest" refers to any nucleotide sequence, the manipulation of which may be deemed desirable for any reason (e.g., treat disease, confer improved qualities, etc.), by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (e.g., reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factors, etc.), and noncoding regulatory sequences which do not encode an mRNA or protein product (e.g., promoter sequence, polyadenylation sequence, termination sequence, enhancer sequence, etc.).

5

10

15

20 -

25

30

As used herein, the terms "complementarity," or "complementary" are used in reference to nucleotide sequences related by the base-pairing rules. For example, the sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

A "complement" of a nucleic acid sequence as used herein refers to a nucleotide sequence whose nucleic acids show total complementarity to the nucleic acids of the nucleic acid sequence.

The term "homology" when used in relation to nucleic acids refers to a degree of complementarity. There may be partial homology (i.e., partial identity) or complete homology (i.e., complete identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid sequence and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe (i.e., an oligonucleotide which is capable of hybridizing to another oligonucleotide of interest) will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence to a target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific

binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

5

10

15

20

25

30

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize to the single-stranded nucleic acid sequence under conditions of low stringency as described *infra*.

The term "hybridization" as used herein includes "any process by which a strand of nucleic acid joins with a complementary strand through base pairing." [Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY]. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the  $T_m$  of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein, the term " $T_m$ " is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the  $T_m$  of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the  $T_m$  value may be calculated by the equation:  $T_m = 81.5 + 0.41(\% G + C)$ , when a nucleic acid is in aqueous solution at 1 M NaCl [see e.g., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985)]. Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of  $T_m$ .

Low stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH<sub>2</sub>PO<sub>4</sub>•H<sub>2</sub>O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 1% SDS, 5X Denhardt's reagent [50X Denhardt's contains the following per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml

denatured salmon sperm DNA followed by washing in a solution comprising 0.2X SSPE, and 0.1% SDS at room temperature when a DNA probe of about 100 to about 1000 nucleotides in length is employed.

5

10

15

20

25

30

High stringency conditions when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 68°C in a solution consisting of 5X SSPE, 1% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, and 0.1% SDS at 68°C when a probe of about 100 to about 1000 nucleotides in length is employed.

The term "equivalent" when made in reference to a hybridization condition as it relates to a hybridization condition of interest means that the hybridization condition and the hybridization condition of interest result in hybridization of nucleic acid sequences which have the same range of percent (%) homology. For example, if a hybridization condition of interest results in hybridization of a first nucleic acid sequence with other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence, then another hybridization condition is said to be equivalent to the hybridization condition of interest if this other hybridization condition also results in hybridization of the first nucleic acid sequence with other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence.

When used in reference to nucleic acid hybridization the art knows well that numerous equivalent conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency hybridization different from, but equivalent to, the above-listed conditions.

Those skilled in the art know that whereas higher stringencies may be preferred to reduce or eliminate non-specific binding of the nucleotide sequence of SEQ ID NOs:1 or 3 with other nucleic acid sequences, lower stringencies may be preferred to detect a larger number of nucleic acid sequences having different homologies to the nucleotide sequence of SEQ ID NOs:1 and 3.

As used herein, the terms "regulatory element" and "regulatory sequence" interchangeably refer to a nucleotide sequence which does not encode RNA or a protein and which controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element which facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, etc. In contrast, the term "regulatory gene" refers to a DNA sequence which encodes RNA or a protein (e.g., transcription factor) that controls the expression of other genes.

5

10

15

20

25

30

Regulatory elements may be tissue specific or cell specific. The term "tissue specific" as it applies to a regulatory element refers to a regulatory element that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (e.g., liver) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue (e.g., lung).

Tissue specificity of a regulatory element may be evaluated by, for example, operably linking a reporter gene to a promoter sequence (which is not tissue-specific) and to the regulatory element to generate a reporter construct, introducing the reporter construct into the genome of an animal such that the reporter construct is integrated into every tissue of the resulting transgenic animal, and detecting the expression of the reporter gene (e.g., detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic animal. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the regulatory element is "specific" for the tissues in which greater levels of expression are detected. Thus, the term "tissue-specific" (e.g., liver-specific) as used herein is a relative term that does not require absolute specificity of expression. In other words, the term "tissue-specific" does not require that one tissue have extremely high levels of expression and another tissue have no expression. It is sufficient that expression is greater in one tissue than another. By contrast, "strict" or "absolute" tissue-specific expression is meant to indicate expression in a single tissue type (e.g., liver) with no detectable expression in other tissues.

The term "cell type specific" as applied to a regulatory element refers to a regulatory element which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide

sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a regulatory element also means a regulatory element capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue.

5

10

15

20

25

30

Cell type specificity of a regulatory element may be assessed using methods well known in the art, e.g., immunohistochemical staining and/or Northern blot analysis. Briefly, for immunohistochemical staining, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is regulated by the regulatory element. A labeled (e.g., peroxidase conjugated) secondary antibody which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (e.g., with avidin/biotin) by microscopy. Briefly, for Northern blot analysis, RNA is isolated from cells and electrophoresed on agarose gels to fractionate the RNA according to size followed by transfer of the RNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized RNA is then probed with a labeled oligo-deoxyribonucleotide probe or DNA probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists.

The term "promoter," "promoter element," or "promoter sequence" as used herein, refers to a DNA sequence which when ligated to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into mRNA. A promoter is typically, though not necessarily, located 5' (i.e., upstream) of a nucleotide sequence of interest whose transcription into mRNA it controls, and provides a site for specific binding by RNA polymerase and other transcription factors for initiation of transcription.

Promoters may be constitutive or regulatable. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing transcription of an operably linked nucleic acid sequence in the absence of a stimulus (e.g., heat shock, chemicals, etc.). In contrast, a "regulatable" promoter is one which is capable of directing a level of transcription of an operably linked nucleic acid sequence in the presence of a stimulus (e.g., heat shock, chemicals, etc.) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

The terms "essentially consisting of" and "consisting essentially of" are equivalent terms, and when in reference to a nucleic acid sequence they are intended to refer to nucleotide sequences which contain from 50% to 100% of the nucleic acid bases which are present in the nucleic acid sequence, in which the arrangement of these nucleic acid bases with respect to each other in the nucleotide sequences is the same as their arrangement in the nucleic acid sequence, and in which the biological activity of the nucleotide sequences is from 50% to 100%, more preferably from 75% to 100%, and most preferably from 90% to 100%, of the biological activity of the nucleic acid sequence. To illustrate, the term "a nucleic acid sequence consisting essentially of SEQ ID NO:1" refers to nucleotide sequences which contain from 50% to 100% of the nucleic acid bases which are present in SEQ ID NO:1, in which the arrangement of these nucleic acid bases with respect to each other in the nucleotide sequences is the same as their arrangement in SEQ ID NO:1, and in which the nucleotide sequences exhibit from 50% to 100%, more preferably from 75% to 100%, and most preferably from 90% to 100%, of the age-related regulatory activity, and/or of the liver-specific activity of SEQ ID NO:1.

5

10

15

20

25

30

A "functional homolog" of a nucleotide sequence which is derived from the hFIX gene shown in Figure 8 is defined as a nucleic acid sequence which has more than 50% identity and less than 100% identity with the hFIX-derived nucleotide sequence (i.e., with the entire, or a portion of the, hFIX sequence of Figure 8), and which has age-related regulatory activity and/or liver-specific activity. For example, a functional homolog of SEQ ID NO:33 includes nucleic acid sequences which have more than 50% identity and less than 100% identity with SEQ ID NO:33, and which have age-related regulatory activity and/or liver-specific activity.

A "functional homolog" of a nucleotide sequence which is derived from the hPC sequence shown in Figure 14 is defined as a nucleic acid sequence which has more than 50% identity and less than 100% identity with the hPC-derived nucleotide sequence (*i.e.*, with the entire, or a portion of the, hPC sequence of Figure 14), and which has age-related regulatory activity and/or regulatory activity. For example, a functional homolog of SEQ ID NO:93 includes nucleic acid sequences which have more than 50% identity and less than 100% identity with SEQ ID NO:93, and which have age-related regulatory activity and/or regulatory activity.

### **DESCRIPTION OF THE INVENTION**

5

10

15

20

25

The invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest. In on embodiment, the invention provides nucleic acid sequences which regulate expression of a nucleotide sequence of interest in an age-related manner. Yet more particularly, the exemplary age-regulatory element 5' (AE5') has been discovered to regulate stable gene expression over time in vivo, while the exemplary ageregulatory element 3' (AE3') has been discovered to regulate increased gene expression over time in vivo. In another embodiment, the invention provides nucleic acid sequences (e.g., AE5') which direct liver-specific expression of a gene of interest. In yet another embodiment, the invention provides transgenic animals which harbor the nucleic acid sequences provided herein and which expres a nucleotide sequence of interest in an agerelated and/or liver-specific manner. The nucleic acid sequences provided herein are useful in, for example, identifying and isolating functional homologs of AE5' and AE3', and amplifying at least a portion of AE5' and AE3'. Importantly, the nucleic acid sequences of the invention are also useful in age-related expression and/or liver-specific expression of a nucleotide sequence of interest in an animal, in gene therapy, and in reducing expression of factor IX in an animal.

In addition to regulatory sequences which are derived from the human factor IX gene (e.g., AE5', AE3', and AE3''), the invention further provides sequences which are derived from the human protein C (hPC) gene and which are characterized by age-related regulatory activity and regulatory activity.

The invention is further discussed under (A) Regulatory Nucleic Acid Sequences, (B) Using Probes To Identify And Isolate Homologs Of AE5', AE3', and Of hPC-Derived Regulatory Sequences, (C) Using Primers to Amplify At Least A Portion Of AE5', AE3', and Of hPC-Derived Regulatory Sequences, (D) Methods For Regulating Gene Expression, (E) Gene Therapy, and (F) Reducing Expression Of Factor IX In An Animal.

### A. Regulatory Nucleic Acid Sequences

5

10

15

20

25

30

The invention provides regulatory sequences which are derived from the hFIX and hPC genes.

### i. Regulatory Nucleic Acid Sequences From The hFIX Gene

The regulatory nucleic acid sequences of the invention and their surprising properties in regulating gene expression were discovered during the inventor's investigation of the mechanisms underlying age-associated regulation of the human factor IX, which is involved in blood coagulation. Blood coagulation plays a critical role not only in homeostasis, but also in many physiological and pathological conditions [Saito in Disorders of Hemostasis, O.D. Ratnoff and C.D. Forbes, Eds., Sauders, Philadelphia, ed. 2 (1991), pp. 18-47; Kurachi et al. (1993) Blood Coagul. Fibrinol. 4:953-974]. Blood coagulation potential in humans as well as in other mammals reaches the young adult level around the age of weaning [Yao et al. (1991) Thromb. Haemost. 65:52-58; Andrew et al. (1992) Blood 80:1998-2005; Andrew et al. Blood (1987) 70:165-172; Andrew et al. (1988) Blood 72:1651-1657]. This is followed by a gradual increase in coagulation potential during young adulthood, and an almost two-fold increase by old age [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi et al. (1996) Thromb. Haemost. 76:965-969]. This age-associated increase in coagulation potential takes place in healthy centenarians [Marie et al. (1995)] Blood 85:3144-3149], indicating that the increase is a normal phenomenon associated with aging.

It is the inventors' consideration that this increase in coagulation potential may make a crucial contribution to the development and progression of age-associated diseases such as cardiovascular and thrombotic disorders [Conlan et al. (1993) The Atherosclerosis Risk in Communities (ARIC) Study 70:380-385; Balleisen et al. (1985) Thromb. Haemost. 54:475-479; Rode et al. (1996) Nat. Med. 2:293-298; Woodward et al. (1997) Brit. J. Haemat. 97:785-797]. The inventors' consideration was based on the observation that this increase in blood coagulation potential coincides with plasma level increases of pro-coagulant factors such as factor IX, whereas plasma levels of anti-coagulation factors (such as antithrombin III and protein C) or of factors involved in fibrinolysis are only marginally affected [Conlan et al. (1994) The Atherosclerosis Risk in Committees (ARIC) Study 72:551-556; Lowe et al. (1997) Brit. J. Haemat. 97:775-784]. These facts strongly suggested to the inventors that the

observed increase in blood coagulation activity with advancing age is due to regulated events. Plasma levels of each protein factor involved in blood coagulation, fibrinolysis and their regulatory systems are presumably determined by the balance of the many processes involved. At present, little is known about why an advancing age-associated increase in blood coagulation activity exists, or what molecular mechanisms are involved in age-dependent regulation (homeostasis) of blood coagulation [Finch in *Longevity, Senescence, and the Genome*, The University of Chicago Press, Chicago, 1990].

5

10

15

20

25

30

Blood coagulation factor IX (FIX) occupies a key position in the blood coagulation cascade where the intrinsic and extrinsic pathways merge [Saito in *Disorders of Hemostasis*, O.D. Ratnoff and C.D. Forbes, Eds., Sauders, Philadelphia, ed. 2 (1991), pp. 18-47; Kurachi et al. (1993) Blood Coagul. Fibrinol. 4:953-974]. FIX is synthesized in the liver with strict tissue-specificity, and its deficiency results in the bleeding disorder hemophilia B. In normal humans the plasma activity and protein concentration levels of human FIX (hFIX) increase with advancing age [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi et al. (1996) Thromb. Haemost. 76:965-969]. Mouse FIX (mFIX) plasma activity also increases with age in a manner similar to hFIX, and is directly correlated with an increase in liver mFIX messenger RNA (mRNA) level [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi et al. (1996) Thromb. Haemost. 76:965-969]. However, nothing else is known about the molecular mechanisms underlying such an increase. In investigating the basic molecular mechanisms responsible for age-associated regulation of hFIX, the inventors discovered the nucleotide sequences which regulate age-associated expression, and which direct liver-specific expression, of the exemplary hFIX gene.

The discovery of the invention sequences was made possible, in part, by the inventors' use of the hFIX promoter in combination with the coding sequence for hFIX instead of with the coding sequence for commonly used reporter proteins. The discovery of the surprising functions of the nucleotide sequences provided herein was also made possible by the inventors' use of longitudinal *in vivo* analyses, rather than of *in vitro* analyses. In particular, the inventors' earlier studies used reporter genes (including bacterial β-galactosidase and chloramphenicol acetyltransferase [CAT]) which are heterologous to the factor IX promoter. In these earlier studies, the factor IX promoter showed only very weak expression activity *in vitro* [Kurachi et al. (1995) J. Biol. Chem. 270:5276-5281]. Use of such heterologous reporter genes made it impossible to reliably and quantitatively perform

longitudinal analyses of transgene expression in animals. The inventors unexpectedly observed that the use of hFIX minigene expression vectors which contained the hFIX promoter and its homologous hFIX gene were capable of producing high level plasma hFIX in vivo. This unexpected observation not only solved the problems associated with the use of genes which are heterologous to the hFIX promoter by providing a reliable animal assay system, but also provided multiple unexpected critical insights into the regulatory mechanisms of the hFIX gene, including the determination of nucleotide sequences which regulate the stability and age-related increased expression of the exemplary hFIX gene.

5

10

15

20

25

30

The present invention provides the 32-nucleotide nucleic acid sequence 5'-agccatt cagtcgagga aggatagggt ggtat-3' (SEQ ID NO:1) of AE5' which corresponds to the sequence from 2164 to 2195 of the hFIX gene deposited in GenBank as accession number K02402, and which corresponds to the sequence from -802 to -771 of Figure 8 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30.

The present invention also provides the 1273-nucleotide nucleic acid sequence (SEQ ID NO:3) (Figure 13) of AE3' which corresponds to the sequence from 34,383 to 35,655 of GenBank accession number K02402, and which corresponds to the sequence from 31,418 to 32,690 of Figure 8 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30.

The terms "age-related regulatory activity" and "age-related activity" when made in reference to a nucleic acid sequence refer to the ability of the nucleic acid sequence to alter in an age-related manner (e.g., increase over a period of time) the level of transcription into mRNA and/or the synthesis of a polypeptide encoded by a nucleotide sequence of interest which is operably linked to a promoter sequence as compared to the level of transcription into mRNA of the nucleotide sequence of interest which is operably linked to the promoter sequence in the absence of the nucleic acid sequence which has age-related regulatory activity. An "age regulatory sequence" is herein used to refer to a nucleic acid sequence which has age-related regulatory activity.

To illustrate, where expression levels of a gene of interest decrease over a period of time, a nucleic acid sequence is said to have age-related regulatory activity if (when operably linked to the gene of interest) it results in (a) a smaller decrease in expression levels of the gene over the same period of time as compared to the decrease in expression levels in the absence of the nucleic acid sequence, (b) relatively constant (i.e., unchanged) expression

levels over the same period of time, or (c) increased expression levels over the same period of time.

5

10

15

20

25

30

The terms "operably linked," "in operable combination," and "in operable order" as used herein refer to the linkage of nucleic acid sequences such that they perform their intended function. For example, operably linking a promoter sequence to a nucleotide sequence of interest refers to linking the promoter sequence and the nucleotide sequence of interest in a manner such that the promoter sequence is capable of directing the transcription of the nucleotide sequence of interest and/or the synthesis of a polypeptide encoded by the nucleotide sequence of interest. Similarly, operably linking a nucleic acid sequence having age-related regulatory activity to a promoter sequence and to a nucleotide sequence of interest means linking the nucleic acid sequence having age-related regulatory activity, the promoter sequence and the nucleotide sequence of interest in a manner such that the nucleic acid sequence having age-related regulatory activity is capable of altering over a period of time the level of transcription into mRNA of the nucleotide sequence of interest and/or the synthesis of a polypeptide encoded by the nucleotide sequence of interest.

Methods for determining age-related regulatory activity of a candidate nucleic acid sequence, given the teachings of the present specification, are within the ordinary skill in the art and are exemplified by the methods disclosed herein. For example, a test vector is constructed in which the candidate nucleic acid sequence is linked upstream or downstream of a promoter sequence which is operably linked to a nucleotide sequence of interest (e.g., Example 1). A control vector which is similar to the test vector but which lacks the candidate nucleic acid sequence is also constructed. The test vector and control vector are separately introduced into a host cell. It is preferred that the host cell (e.g., fertilized egg) be capable of generating a transgenic multicellular organism, e.g., a transgenic mouse (e.g., Example 3) and that transgenic multicellular organisms are generated. Longitudinal analyses of the expression of mRNA which is encoded by the nucleotide sequence of interest (e.g., by Northern blot hybridization) over a period of time in, and preferably over the entire life span of, founders and successive generations of the transgenic multicellular organism are carried out (e.g., Example 3). The detection in any tissue of mRNA and/or protein levels which are encoded by the nucleotide sequence of interest and which are greater in transgenic animals harboring the test vector as compared to the mRNA and/or protein levels in transgenic

animals harboring the control vector at least one point in time indicates that the candidate nucleic acid sequence has age-related regulatory activity.

For example, evidence provided herein shows the surprising result that AE5' (SEQ ID NO:1) alone has age-related regulatory activity in that AE5' stabilizes hFIX mRNA whereby hFIX mRNA levels are essentially unchanged at different time points over the entire life span of transgenic animals (Figure 4, A, C and E) as compared to the declining hFIX mRNA levels in transgenic animals which harbor vectors that lack AE5' (Figures 2A and 2E). The age-related regulatory activity of AE5' was observed regardless whether AE5' was placed upstream (Figure 4A) or downstream (Figure 4E) of the promoter sequence in the expression construct.

5

10

15

20

25

30

Furthermore, data provided herein demonstrates the unexpected result that AE3' (SEO ID NO:3) alone has age-related regulatory activity in that AE3' increases hFIX mRNA at several time points during the life of transgenic animals (Figure 2B) relative to the hFIX mRNA levels at the same time points in transgenic animals harboring vectors that lack AE3' (Figure 2A). AE3' substantially increased the steady state hFIX mRNA levels (Figure 5). This result which was observed in vivo was surprising in part because AE3' exhibited weak down regulatory effects on hFIX production in vitro. Without limiting the invention to any particular mechanism, these results suggest that AE3' functions by increasing hFIX mRNA stability which directly correlates with an increase in the hFIX protein level in the circulation. Also without intending to limit the invention to any particular theory, it is the inventors' consideration that the age-related regulatory activity of AE3' is due to the s1 structure-forming dinucleotide repeats present in the 3'UTR; the s1 region is the 103 bp sequence (SEQ ID NO:61) from nt 32,141 through nt 32,243 of Figure 8. This consideration is based on the inventors' observation that dinucleotide repeats, such as (AT), of the 3' UTR of various genes, can form sl structures in mRNA, which upon binding specific proteins are known to modulate mRNA stability, mostly to a less stable state [Ross (1995) Microbiol. Rev. 59:423-450].

Importantly, the invention demonstrates the surprising synergistic action of AE5' and AE3' which together result in hFIX mRNA levels which not only are greater at each time point tested over the life span of transgenic animals (Figures 4 B and D) as compared to hFIX mRNA levels in transgenic animals harboring vectors that lack both AE5' and AE3', but also that the profile of increased human FIX mRNA levels over the life span of

transgenic mouse recapitulates the profile of increased mouse FIX mRNA levels as a wildtype mouse ages.

Data presented herein further demonstrate that the age-related regulatory activity of AE5' alone, of AE3' alone, and of the combination of AE5' and AE3' is independent of the level of expression of the transgenes harboring them, sex, generation or zygosity status of the transgenic animals.

5

. 10

15

20

25

30

The present invention is not limited to SEQ ID NOs:1 and 3 but specifically contemplates portions thereof. As used herein the term "portion" when made in reference to a nucleic acid sequence refers to a fragment of that sequence. The fragment may range in size from five (5) contiguous nucleotide residues to the entire nucleic acid sequence minus one nucleic acid residue. Thus, a nucleic acid sequence comprising "at least a portion of" a nucleotide sequence comprises from five (5) contiguous nucleotide residues of the nucleotide sequence to the entire nucleotide sequence.

In a preferred embodiment, portions of SEQ ID NO:1 contemplated to be within the scope of the invention include, but are not limited to, the 7-nucleotide nucleic acid sequence of the polyomavirus enhance activator 3 (PEA-3) (5'-GAGGAAG-3') (SEQ ID NO:2) which corresponds to the sequence from 2176 to 2182 of GenBank accession number K02402, and which corresponds to the sequence from -790 to -784 of GenBank accession number K02402 when in relation to the hFIX start codon (ATG) in which the adenine is designated as position +30. A nucleotide sequence [5'-CAGGAAG-3' (SEQ ID NO:40)] which is homologous to the invention's PEA-3 nucleotide sequence was initially reported in the art as a polyoma virus enhancer, and was reported to be involved in the regulation of expression of various genes (e.g., collagen gene and c-fos) in several tissues [Martin et al. (1988) Proc. Natl. Acad. Sci. 85:5839-5843; Xin et al. (1992) Genes & Develop. 6:481-496; Chotteau-Lelievre et al. (1997) Oncogene 15:937-952; Gutman and Wasylyk (1990) EMBO J. 9:2241-2246]. However, the PEA-3 protein sequence [or PEA-3 related protein(s)] which binds to nucleotide sequences which are homologous to the invention's PEA-3 nucleotide sequence has not been reported to be either liver-specific or enriched in the liver.

Other portions of SEQ ID NO:1 included within the scope of the invention include, for example, SEQ ID NO:33 [5'-tcgaggaagga-3'], SEQ ID NO:34 [5'-agtcgaggaaggata-3'], SEQ ID NO:35 [5'-tcagtcgaggaaggatagg-3'], SEQ ID NO:36 [5'-attcagtcgaggaaggatagggt-3'], SEQ ID NO:37 [5'-ccattcagtcgaggaaggatagggtgg-3'], and

SEQ ID NO:38 [5'-gccattcagtcgaggaaggatagggtggta-3'], all of which include the PEA-3 nucleotide sequence.

5

10

15

20

25

30

In a preferred embodiment, portions of SEO ID NO:3 contemplated to be within the scope of the invention include, but are not limited to, SEQ ID NO:51 [5'-TTATTTTATATATATATATATATATAAAATA-3'], SEQ ID NO:52 [5'-TAT AATATA-3'], SEQ ID NO:53 [5'-CAATATAAATATAG-3'], SEQ ID NO:54 [5'-combination of SEQ ID NOs:51 and 52, i.e., SEQ ID NO:55 [5'-TTATTTTATA TATATATATATATAAAATATATAATATA-3'], the combination of SEQ ID NOs:52 and 53, i.e., SEQ ID NO:56 [5'-TATAATATACAATATAAATATAG-3'], the combination of SEQ ID NOs:53 and 54, i.e., SEQ ID NO:57 [5'-CAATATAAAT combination of SEQ ID NOs:51, 52, 53, and 54, i.e., SEQ ID NO:58 [5'-TTAT TTTATATATATATATATATAAAATATATAATATACAATATAAATATAGTGT GTGTGTATGCGTGTGTAGACACACACACACACACACATA-3'], the 723 bp sequence (SEQ ID NO:59) from nt 31,418 through nt 32,140 of Figure 8, the 447 bp sequence (SEQ ID NO:60) from nt 32,244 through nt 32,690 of Figure 8, and the 103 bp sequence (SEQ ID NO:61) (i.e., the s1 region of the 3' UTR) from nt 32,141 through nt 32,243 of Figure 8.

The nucleotide sequence of portions of SEQ ID NOs:1 and 3 which exhibit agerelated regulatory activity may be determined using methods known in the art, e.g., using deletion constructs (e.g., see Yang et al. (1998) J. Biol. Chem. 273:891-897). Briefly, several expression plasmids are constructed to contain a reporter gene under the control of a promoter and of different candidate nucleotide sequences which are obtained either by restriction enzyme deletion of internal sequences in SEQ ID NOs:1 and 3, restriction enzyme truncation of sequences at the 5' and/or 3' end of SEQ ID NOs:1 and 3, by the introduction of single nucleic acid base changes by PCR into SEQ ID NOs:1 and 3, or by chemical synthesis. The gene-related regulatory activity of the different constructs is determined as described supra in order to determine whether the candidate nucleotide sequence exhibits age-related regulatory activity.

The sequences of the present invention are not limited to SEQ ID NOs:1 and 3 and portions thereof, but also include homologs of SEQ ID NOs:1 and 3, and homologs of portions thereof. Homologs of SEQ ID NOs:1 and 3, and of portions thereof, include, but

are not limited to, nucleotide sequences having deletions, insertions or substitutions of different nucleotides or nucleotide analogs as compared to SEQ ID NOs:1 and 3, and of portions thereof, respectively. Such homologs may be produced using methods well known in the art.

5

10

15

20

25

30

A "homolog" of SEQ ID NO:1 is defined as a nucleotide sequence having more than 63% identity and less than 100% identity with SEQ ID NO:1. Homologs of SEQ ID NO:1 are exemplified, but not limited to, SEQ ID NO:66 (5'-acccatt cagtcgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the G at nt 2,165 is replaced with a C; SEQ ID NO:67 (5'-agccatt gagtcgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the C at nt 2,171 is replaced with a G; SEO ID NO:68 (5'agccatt cagacgagga aggatagggt ggtat-3') is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the T at nt 2,174 is replaced with a A; SEQ ID NO:69 (5'-agccatt cagtcgagga aggatagggt ggttt-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that the A at nt 2,194 is replaced with a T; SEQ ID NO:70 (5'-agccatt cagtcgagga tcccaagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that AGGGT beginning at nt 2,186 is replaced with TCCCA; SEQ ID NO:71 (5'-agccatt cagtcgagga aggatagggcctaat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that TGGT beginning at nt 2,190 is replaced with CCTA; SEO ID NO:72 (5'-agaccatt cagtegagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that a A is inserted after nt 2,165; SEQ ID NO:73 (5'-agccatt cagtcgagga aggatagcggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that a C is inserted after nt 2,187; SEQ ID NO:74 (5'-agccatt cagtcgagga aggataat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that GGGTGGT beginning at nt 12,187 is deleted; and SEQ ID NO:75 (5'-agccatt cgagga aggatagggt ggtat-3') which is the sequence from nt 2,164 to nt 2,195 of GenBank accession number k02402, except that CAGT beginning at nt 2,171 is deleted.

A "homolog" of SEQ ID NO:2 is defined as a nucleotide sequence having more than 75% identity and less than 100% identity with SEQ ID NO:2. Homologs of SEQ ID NO:2 include, for example, GAGGATG (SEQ ID NO:39), CAGGAAG (SEQ ID NO:40),

CAGGATG (SEQ ID NO:41), GTGGAAG (SEQ ID NO:62), GTGGATG (SEQ ID NO:63), CTGGAAG (SEQ ID NO:64), CTGGATG (SEQ ID NO:65), and CAGGAAG (SEQ ID NO:84).

5

10

15

20

25

30

A "homolog" of SEQ ID NO:3 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:3. Homologs of SEQ ID NO:3 are exemplified, but not limited to, SEQ ID NOs:76-83 shown in Figure 13. Specifically, SEQ ID NO:76 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the C at nt 34,390 has been replaced with a G. SEQ ID NO:77 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the T at nt 34,649 has been replaced with a A. SEQ ID NO:78 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the GC beginning at nt 34,959 has been replaced with a CG. SEQ ID NO:79 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the CATG beginning at nt 35,501 has been replaced with a GTAC. SEQ ID NO:80 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that TT is inserted after the A at nt 34,681. SEQ ID NO:81 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that TGC is inserted after the C at nt 35,581. SEQ ID NO:82 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that A at nt 35,636 is deleted. SEQ ID NO:83 is the sequence from nt 34,383 to nt 35,655 of GenBank accession number k02402, except that the G at nt 34,383 is deleted.

A "homolog" of SEQ ID NO:59 is defined as a nucleotide sequence having less than 100% and more than 62% identity with SEQ ID NO:59.

A "homolog" of SEQ ID NO:60 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:60.

A "homolog" of SEQ ID NO:61 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:61.

Homologs of a portion of SEQ ID NO:1 are exemplified by homologs of the PEA-3 nucleotide sequence (SEQ ID NO:2), which include, for example, GAGGATG (SEQ ID NO:39), CAGGAAG (SEQ ID NO:40), CAGGATG (SEQ ID NO:41), GTGGAAG (SEQ ID NO:62), GTGGATG (SEQ ID NO:63), CTGGAAG (SEQ ID NO:64), CTGGATG (SEQ ID NO:65), and CAGGAAG (SEQ ID NO:84).

The present invention also contemplates functioning or functional homologs of SEQ ID NO:1, of portions of SEQ ID NO:1 (e.g., functional portions of SEQ ID NOs:2, and 33-38), of SEQ ID NO:3, and of portions of SEQ ID NO:3 (e.g., functional portions of SEQ ID NOs:51-61).

A "functional homolog" of SEQ ID NO:1 is defined as a nucleotide sequence having more than 63% identity and less than 100% identity with SEQ ID NO:1, and which has agerelated regulatory activity. Alternatively, a functional homolog of SEQ ID NO:1 is a nucleotide sequence having more than 63% identity and less than 100% identity with SEQ ID NO:1, and having liver-specific activity.

5

10

15

20

25

30

A "functional homolog" of SEQ ID NO:2 is defined as a nucleotide sequence having more than 75% identity and less than 100% identity with SEQ ID NO:2, and which has agerelated regulatory activity. Alternatively, a functional homolog of SEQ ID NO:2 is a nucleotide sequence having more than 75% identity and less than 100% identity with SEQ ID NO:2, and having liver-specific activity.

A "functional homolog" of SEQ ID NO:3 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:3, and which has age-related regulatory activity.

A "functional homolog" of SEQ ID NO:59 is defined as a nucleotide sequence having less than 100% and more than 62% identity with SEQ ID NO:59, and which has age-related regulatory activity.

A "functional homolog" of SEQ ID NO:60 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:60, and which has age-related regulatory activity.

A "functional homolog" of SEQ ID NO:61 is defined as a nucleotide sequence having less than 100% and more than 60% identity with SEQ ID NO:61, and which has age-related regulatory activity.

The present invention is not limited to sense molecules of SEQ ID NOs:1 and 3 but contemplates within its scope antisense molecules comprising a nucleic acid sequence complementary to at least a portion (e.g., a portion greater than 10 nucleotide bases in length and more preferably greater than 100 nucleotide bases in length) of the nucleotide sequence of SEQ ID NOs:1 and 3. These antisense molecules find use in, for example, reducing or

preventing expression of a gene (e.g. hFIX) whose expression is regulated by SEQ ID NOs:1 and 3.

5

10

15

20

25

30

Data presented herein demonstrates the universality of the regulatory function of portions of AE3' in that AE3'' has been successfully used to regulate expression of the heterologous hPC gene in an age-related manner. In particular, transgenic animals containing the -1462hPCm1 construct expressed age-stable levels of human protein C, *i.e.*, expressed relatively constant levels of human protein C at different time points during the life span of the transgenic animals (Example 12, Figure 17A). In direct contrast, the presence of AE5' and AE3'' sequences resulted in increased expression levels of human protein C over time (Figure 17C). These results confirm the universality of the function of the AE3'' portion of AE3' in regulating expression of operably linked genes in an age-related manner.

The invention also contemplates portions of AE3". These portions include, but are not limited to, SEQ ID NOs:94-144 wherein SEQ ID NO:94 is 5'-tgggg gaaaagtttc tttcagagag ttaagttatt ttatatata aatatatata taaaatatat aatatacaat ataaaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:95 is 5'-gggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:96 is 5'-ggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaaatatat aatatacaat ataaaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:97 is 5'-gg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:98 is 5'-g gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:98 is 5'-g gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:99 is 5'-

5

10

15

20

25

30

gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:100 is 5'-aaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:101 is 5'-aaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEO ID NO:102 is 5'-aagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEO ID NO:103 is 5'-agttic titcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:104 is 5'-gtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:105 is 5'-tttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:106 is 5'-ttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:107 is 5'-tc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgegtg tgtgtagaca cacaegcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:108 is 5'-c tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:109 is 5'-tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:110 is 5'-ttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:111 is 5'-tcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:112 is 5'-cagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEO ID NO:113 is 5'-agagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEO ID NO:114 is 5'-gagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:115 is 5'-agag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata

5

10

15

20

25

30

cacacatata atggaagcaa taagccat-3'; SEQ ID NO:116 is 5'-gag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:117 is 5'-ag ttaagttatt ttatatata aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:118 is 5'-g ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:119 is 5'-ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:120 is 5'-taagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:121 is 5'-aagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:122 is 5'-agttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3': SEO ID NO:123 is 5'-gttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:124 is 5'-ttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccat-3'; SEQ ID NO:125 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagcca-3'; SEQ ID NO:126 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagcc-3'; SEQ ID NO:127 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagc-3'; SEQ ID NO:128 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taag-3'; SEQ ID NO:129 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taa-3'; SEQ ID NO:130 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa ta-3'; SEQ ID NO:131 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa t-3'; SEQ ID NO:132 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata



5

10

15

20

25

30

atggaagcaa -3'; SEQ ID NO:133 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagca-3'; SEQ ID NO:134 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagc-3'; SEQ ID NO:135 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaag-3'; SEQ ID NO:136 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaa-3'; SEQ ID NO:137 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atgga-3'; SEQ ID NO:138 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atgg-3'; SEQ ID NO:139 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatata atg-3'; SEQ ID NO:140 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg igtatgcgtg tgtgtagaca cacacgcata cacacatata at-3'; SEQ ID NO:141 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata a-3'; SEQ ID NO:142 is 5'-ttgggg gaaaagtttc tticagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata-3'; SEQ ID NO:143 is 5'-ttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatata taaaatatat aatatacaat ataaatatat agtgtgtgt tgtatgcgtg tgtgtagaca cacacgcata cacacatat-3'; and SEQ ID NO:144 is 5'-ttgggg gaaaagttic tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacata-3'.

The nucleotide sequence of SEQ ID NOs:1 and 3, portions, homologs and antisense sequences thereof may be synthesized by synthetic chemistry techniques which are commercially available and well known in the art [see Caruthers MH et al., (1980) Nuc. Acids Res. Symp. Ser. 215-223; Horn T. et al., (1980) Nuc. Acids Res. Symp. Ser. 225-232]. Additionally, fragments of SEQ ID NOs:1 and 3 can be made by treatment of SEQ ID NOs:1 and 3 with restriction enzymes followed by purification of the fragments by gel electrophoresis. Alternatively, sequences may also be produced using the polymerase chain reaction (PCR) as described by Mullis [U.S. Patent Nos. 4,683,195, 4,683,202 and 4,965,188, all of which are hereby incorporated by reference]. SEQ ID NOs:1 and 3, portions, homologs and antisense sequences thereof may be ligated to each other or to heterologous nucleic acid sequences using methods well known in the art.

5

10

15

20

25

30

The nucleotide sequence of synthesized sequences may be confirmed using commercially available kits as well as using methods well known in the art which utilize enzymes such as the Klenow fragment of DNA polymerase I, Sequenase<sup>®</sup>, *Taq* DNA polymerase, or thermostable T7 polymerase. Capillary electrophoresis may also be used to analyze the size and confirm the nucleotide sequence of the products of nucleic acid synthesis, restriction enzyme digestion or PCR amplification.

It is readily appreciated by those in the art that the sequences of the present invention may be used in a variety of ways. For example, the nucleic acid sequences of the invention and portions thereof can be used as probes for the detection and isolation of functional homologs of AE5' and AE3', amplification of homologous nucleotide sequences, age-related and/or liver-specific expression of a nucleotide sequence of interest in an animal, gene therapy, and reducing factor IX levels in an animal.

ii. Regulatory Nucleic Acid Sequences From The hPC Gene

The invention provides regulatory nucleic acid sequences which are derived from the hPC gene. The presence of these sequences and their surprising properties was fortuitously discovered by the inventors during their investigation of the universality of the function of regulatory sequences from the hFIX gene. In particular, when using what they believed would be "control" constructs which contained different portions of the sequence upstream of nt +1 in the hPC gene, the inventors discovered that whereas transgenic animals containing the -1462hPCm1 construct exhibited relatively constant and relatively high levels (from

about 100 to about 3000 ng/ml) of human protein C over time (Figure 17A), in dramatic contrast, transgenic animals containing the -82hPCml construct exhibited relatively low levels (from about 5 to about 40 ng/ml at 1 month of age) (Figure 17B) of human protein C which declined at a precipitous rate over time (Example 12). Indeed, by the age of 5 months, human protein C levels were undetectable in all transgenic animals harboring the -82hPCml construct. These results demonstrated to the inventors that the nucleotide sequence from nt -1462 to nt -83 of the human protein C gene exhibits age-related regulatory activity (as evidenced by age-stable expression over time) and regulatory activity (as evidenced by the relatively higher levels of expression as compared to the levels in the absence of the nucleotide sequence from nt -1462 to nt -83) of operably linked sequences of interest.

5

10

15

20

25

30

The term "regulatory activity" when made herein in reference to a nucleic acid sequence refers to the ability of the nucleic acid sequence to alter the level of transcription into mRNA and/or the synthesis of a polypeptide encoded by a nucleotide sequence of interest which is operably linked to a promoter sequence, as compared to the level of transcription into mRNA and/or synthesis of the polypeptide encoded by the nucleotide sequence of interest which is operably linked to the promoter sequence in the absence of the nucleic acid sequence which has regulatory activity. In contrast to "age-related regulatory activity," the term "regulatory activity" relates to alteration in the level of transcription or protein synthesis at a single time point (rather than over a period of time) in the life cycle of a cell, tissue, or organism. In a preferred embodiment, the level of transcription into mRNA and/or synthesis of polypeptide is increased from 2-fold to at least 10,000-fold, preferably from 2-fold to 10,000 fold, more preferably from 2-fold to 1,000-fold, and most preferably, from 2-fold to 500-fold, when compared to the level of transcription into mRNA and/or synthesis of the polypeptide encoded by the nucleotide sequence of interest which is operably linked to the promoter sequence in the absence of the nucleic acid sequence which has regulatory activity. Data presented herein demonstrates that the inclusion of the human protein C (hPC) nucleotide sequence from nt -1462 to nt -83 resulted in expression of from about 100 to about 3000 ng/ml of human protein C (Figure 17A) as compared to expression of from about 5 to about 40 ng/ml of human protein C (Figure 17B) in transgenic mice at one-month of age. The increased level of expression was even more dramatic when the transgenic animals were 5 months-old; while the levels of hPC were undetectable in the



absence of the sequence from nt -1462 to nt -83, they were from about 100 to about 3000 ng/ml in the presence of this sequence.

5

10

15

20

25

30

In one embodiment, the invention provides the nucleic acid sequence (SEO ID NO:85) from -1462 to +1 of the hPC gene (Figure 14; 5'-GAATTCTGTA AGCATTTCCT ATGTGTACCT GCCCCTGGGC AAGGTGGGCC TGACTTGTTA GAGTGTTAGA GTTTTACCCT GTTCCTCTAG GAGGGCCTGG TACCACCACA GCCCAGCATG GTGTGGTGCC TCAGCAGGAG GCATCTGGTT ACAATCAACA CAAGCTGTTC CAGCCAATTT AAAGAAACTT CAGGAGGAAT AGGGTTTTAG GAGGGCATGG GGACCCTCCT GCACCCGAAG CCAGGATGTG CCACCAATCA TAAGGAGGCA GGGGCCTCCT TCCGCTGCTC CCTGGGACTC TCTAGGTGTC CGTGGCCTCA GCCCCCTCT GCACACCTGC ATCTTCCTTC TCATCAGCTT CCTCTGCTTT AAGCGTAAAC ATGGATGCCC AGGACCTGGC CTCAATCTTC CGAGTCTGGT ACTTATGGTG TACTGACAGT GTGAGACCCT ACTCCTCTGA TCAATCCCCT GGGTTGGTGA CTTCCCTGTG CAATCAATGG AAGCCAGCGA GGCAGGGTCA CATGCCCCGT TTAGAGGTGC AGACTTGGAG AAGGAACGTG GGCAAGTCTT CCCAGGAACA GGTAGGGCAG GGAGGAAAGG GGGGCATCTC TGGTGCAGCC CGGTTCGGAG CAGGAAGACG CTTAATAAAT GCTGATAGAC TGCAGGACAC AGGCAAAGGT GCTGAGCTGG ACCCTTTATT TCTGCCCTTC TCCCTTCTGG CACCCGGCC AGGAAATTGC TGCAGCCTTT CTGGAATCCC GTTCATTTTT CTTACTGGTC CACAAAAGGG GCCAAATGGA AGCAGCAAGA CCTGAGTTCA AATTAAATCT GCCAACTACC AGCTCAGTGA ATCTGGGCGA GTAACACAAA ACTTGAGTGT CCTTACCTGA AAAATAGAGG TTAGAGGGAT GCTATGTGCC ATTGTGTGT TGTGTTGGGG GTGGGGATTG GGGGTGATTT GTGAGCAATT GGAGGTGAGG GTGGAGCCCA GTGCCCAGCA CCTATGCACT GGGGACCCAA AAAGGAGCAT CTTCTCATGA TTTTATGTAT CAGAAATTGG GATGGCATGT CATTGGGACA GCGTCTTTTT TCTTGTATGG TGGCACATAA ATACATGTGT CTTATAATTA ATGGTATTTT AGATTTGACG AAATATGGAA TATTACCTGT TGTGCTGATC TTGGGCAAAC TATAATATCT CTGGGCAAAA ATGTCCCCAT CTGAAAACA GGGACAACGT TCCTCCCTCA GCCAGCCACT ATGGGGCTAA AATGAGACCA CATCTGTCAA GGGTTTTGCC CTCACCTCCC TCCCTGCTGG ATGGCATCCT TGGTAGGCAG AGGTGGGCTT CGGGCAGAAC AAGCCGTGCT GAGCTAGGAC CAGGAGTGCT AGTGCCACTG TTTGTCTATG GAGAGGGAGG

CCTCAGTGCT GAGGGCCAAG CAAATATTTG TGGTTATGGA TTA-3'). This sequence was successfully used to express hPC in transgenic mice in an age-related manner (i.e., at relatively constant levels over time; Figure 17A), and at relatively high levels (e.g., as compared to expression levels in the presence of sequences from nt -1462 to nt -82; Figure 17B).

5

10

15

20

25

30

The invention also provides the nucleotide sequence (SEQ ID NO:92) from -1462 to -83 of the hPC gene (Figure 14). The finding that SEQ ID NO:92 exhibited age-related regulatory activity and regulatory activity as discussed in Example 11, infra, was surprising because it was contrary to the results which had previously been reported by Miao et al. [Miao et al. (1996) J. Biol. Chem. 16:9587-9594] when using a heterologous reporter gene, chloramphenicol acetyltranferase (CAT), under the transcriptional control of varying lengths of the protein C 5'-end sequences. In particular, Miao et al. found that construct pPC-1528 which contained nucleotides -1462 to +1 resulted in substantially reduced CAT in vitro activity as compared to construct pPC-82-66 which contained nucleotides -82 to +1. From this, Miao et al: concluded that there is an element with silencer activity in the region between -1462 and -82. In contrast, data presented herein demonstrates that nucleotides -1462 to +1 and -82 to +1 resulted in similar in vitro activities of hPC (Example 11, Figure 16). Indeed, the inventors' data disclosed in relation to hPC's age-related regulatory sequences which are upstream of the hPC coding sequences when used to regulate expression of hPC is consistent with the inventor's observation (Figure 1) in relation to hFIX's agerelated regulatory AE5' sequences when used to regulate expression of hFIX. In particular, the inventors have observed that hFIX's AE3' and AE3" sequences resulted in moderate suppression in in vitro transient expression assays when using hFIX and hPC, respectively (Figures 1 and 16).

Further contemplated to be within the scope of this invention are portions of the SEQ ID NOS:85 and 92. In a particularly preferred embodiment, these portions contain one or both of the first PEA-3 element (SEQ ID NO:89) [7-bp long; 5'-GAGGAAA-3', from -871 to -865 of the hPC gene of Figure 14] and the second PEA-3 element (SEQ ID NO:90) [7-bp long; 5'-CAGGAAG-3', from -832 to -826 of the hPC gene of Figure 14].

Exemplary portions of SEQ ID NOs:85 and 92 which contain both the first and second PEA-3 elements include, but are not limited to, the nucleotide sequence (SEQ ID NO:88) from -1462 to -802 of the hPC gene, which is embodied in plasmid -849hPCm1

(Figure 15, Example 10). Further examples include the nucleotide sequence (SEQ ID NO:145) from -1462 to -804, (SEQ ID NO:146) from -1462 to -805, (SEQ ID NO:147) from -1462 to -806, (SEQ ID NO:148) from -1462 to -807, (SEQ ID NO:149) from -1462 to -808, (SEQ ID NO:150) from -1462 to -809, (SEQ ID NO:151) from -1462 to -810, (SEQ ID NO:152) from -1462 to -811, (SEQ ID NO:153) from -1462 to -812, (SEQ ID NO:154) from -1462 to -813, (SEQ ID NO:155) from -1462 to -814, (SEQ ID NO:156) from -1462 to -815, (SEQ ID NO:157) from -1462 to -816, (SEQ ID NO:158) from -1462 to -817, (SEQ ID NO:159) from -1462 to -818, (SEQ ID NO:160) from -1462 to -819, (SEQ ID NO:161) from -1462 to -820, (SEQ ID NO:162) from -1462 to -821, (SEQ ID NO:163) from -1462 to -822, (SEQ ID NO:164) from -1462 to -823, (SEQ ID NO:165) from -1462 to -824, (SEQ ID NO:166) from -1462 to -825, (SEQ ID NO:167) from -1462 to -826, (SEQ ID NO:168) from -1452 to -803, (SEQ ID NO:169) from -1442 to -803, (SEQ ID NO:170) from -1412 to -803, (SEQ ID NO:171) from -1102 to -803, (SEQ ID NO:172) from -902 to -803, and (SEQ ID NO:173) from -873 to -803.

5

10

15

20

25

30

Portions of SEQ ID NOs:85 and 92 which contain only the first PEA-3 element include the nucleotide sequence (SEQ ID NO:87) from -1462 to -849 of the hPC gene, which is embodied in plasmid -802hPCm1 (Figure 15, Example 10). Additional exemplary portions include, but are not limited to the nucleotide sequence (SEQ ID NO:174) from -1462 to -850, (SEQ ID NO:175) from -1462 to -851, (SEQ ID NO:176) from -1462 to -852, (SEQ ID NO:177) from -1462 to -853, (SEQ ID NO:178) from -1462 to -854, (SEQ ID NO:179) from -1462 to -855, (SEQ ID NO:180) from -1462 to -856, (SEQ ID NO:181) from -1462 to -857, (SEQ ID NO:182) from -1462 to -858, (SEQ ID NO:183) from -1462 to -859, (SEQ ID NO:184) from -1462 to -860, (SEQ ID NO:185) from -1462 to -861, (SEQ ID NO:186) from -1462 to -862, (SEQ ID NO:187) from -1462 to -863, (SEQ ID NO:190) from -1362 to -865, (SEQ ID NO:191) from -1262 to -865, (SEQ ID NO:192) from -1162 to -865, (SEQ ID NO:193) from -1062 to -865, (SEQ ID NO:194) from -962 to -865, and (SEQ ID NO:195) from -872 to -865.

Examples of portions of SEQ ID NOs:85 and 92 which contain only the second PEA-3 element include, but are not limited to, the nucleotide sequence (SEQ ID NO:196) from -863 to -83, (SEQ ID NO:197) from -853 to -83, (SEQ ID NO:198) from -843 to -83, (SEQ ID NO:199) from -833 to -83, (SEQ ID NO:200) from -832 to -83, (SEQ ID NO:201) from

-863 to -183, (SEQ ID NO:202) from -863 to -283, (SEQ ID NO:203) from -863 to -383, (SEQ ID NO:204) from -863 to -483, (SEQ ID NO:205) from -863 to -583, (SEQ ID NO:206) from -863 to -683, (SEQ ID NO:207) from -863 to -783, and (SEQ ID NO:208) from -863 to -826.

In a particularly preferred embodiment, the portion of SEQ ID NO:85 and 92 is selected from the first PEA-3 element (SEQ ID NO:89) and the second PEA-3 element (SEQ ID NO:90). It is the inventor's view that the first and/or second PEA-3 elements within SEQ ID NOs:85 and 92 are responsible for the observed age-related regulatory activity and regulatory activity of SEQ ID NOs:85 and 92.

10

15

20

25

30

5

# B. Using Probes To Identify And Isolate Homologs Of AE5', AE3', and Of hPC-Derived Regulatory Sequences

The invention provided herein is not limited to SEQ ID NO:1, 3, 85 and 92, homologs and portions thereof having age-related regulatory activity, but includes sequences having no age-related regulatory activity (i.e., non-functional homologs and non-functional portions of homologs). The use of such sequences may be desirable, for example, where a portion of SEQ ID NOs:1, 3, 85, and 92 is used as a probe to detect the presence of SEQ ID NOs:1, 3, 85 and 92, respectively, or of portions thereof in a sample.

As used herein, the term "probe" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to a nucleotide sequence of interest. A probe may be single-stranded or double-stranded. It is contemplated that any probe used in the present invention will be labelled with any "reporter molecule," so that it is detectable in any detection system including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, calorimetric, gravimetric, magnetic, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The probes provided herein are useful in the detection, identification and isolation of, for example, sequences such as those listed as SEQ ID NOs:1, 3, 85 and 92 as well as of homologs thereof. Preferred probes are of sufficient length (e.g., from about 9 nucleotides to about 20 nucleotides or more in length) such that high stringency hybridization may be

employed. In one embodiment, probes from 20 to 50 nucleotide bases in length are employed.

# C. Using Primers to Amplify At Least A Portion Of AE5', AE3', and Of hPC-Derived Regulatory Sequences

5

10

15

20

25

30

The invention provided herein is not limited to SEQ ID NOs:1 and 3, homologs and portions thereof having age-related regulatory activity, but includes sequences having no agerelated regulatory activity. This may be desirable, for example, where a portion of the nucleic acid sequences set forth as SEQ ID NOs:1 and 3 is used as a primer for the amplification of nucleic acid sequences by, for example, polymerase chain reactions (PCR) or reverse transcription-polymerase chain reactions (RT-PCR). The term "amplification" is defined as the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction technologies well known in the art [Dieffenbach CW and GS Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY]. As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis disclosed in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,965,188, all of which are hereby incorporated by reference, which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter

"PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are the to be "PCR amplified."

5

10

15

20

25

30

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; and/or incorporation of <sup>32</sup>P-labeled deoxyribonucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any nucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications. Amplified target sequences may be used to obtain segments of DNA (e.g., genes) for the construction of targeting vectors, transgenes, etc.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (i.e., in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long (e.g., from about 9 nucleotides to about 20 nucleotides or more in length) to prime the synthesis of extension products in the presence of the inducing agent. Suitable lengths of the primers may be empirically determined and depend on factors such as temperature, source of primer and the use of the method. In one embodiment, the present invention employs primers from 20 to 50 nucleotide bases in length.

The primers contemplated by the invention are useful in, for example, identifying sequences which are homologous to AE5', AE3', and regulatory sequences derived from hPC, in mammals, yeast, bacteria, and in other organisms.

# D. Methods For Regulating Gene Expression

5

10

15

20

25

30

The present invention provides methods for regulating expression of a nucleotide sequence of interest over a period of time in a cell or multicellular organism. Specifically, gene expression is preferably regulated in a multicellular organism. In one embodiment, expression of a nucleotide sequence of interest is stabilized such that the level of mRNA and/or protein encoded by the nucleotide sequence of interest remains relatively unchanged at different times during the life of the organism. In an alternative embodiment, expression of a nucleotide sequence of interest is increased. Increased expression means that the level of mRNA and/or protein encoded by the nucleotide sequence of interest at a given time point is greater than the level of mRNA and/or protein, respectively, at an earlier time point during the life of the organism or cell. Alternatively, increased expression means that the level of mRNA and/or protein encoded by the nucleotide sequence of interest is greater than the level of mRNA and/or protein encoded by the nucleotide sequence of interest is greater than the level of mRNA and/or protein, respectively, at the same time point in the life of the organism or cell as compared to the level of mRNA and/or protein when expressed in the absence of the sequences of the invention.

In one embodiment, regulating expression of a nucleotide sequence of interest over a period of time is accomplished by introducing into a host cell a vector that contains a nucleotide sequence of interest operably linked to a promoter sequence and to sequences provided herein which have age-related regulatory activity. The transfected host cell is allowed to develop into a transgenic animal in which the nucleotide sequence of interest is expressed in at least one tissue. These steps are further described below for specific embodiments.

#### 1. Expression Constructs

In one embodiment of the methods of the invention for regulating expression of a nucleotide sequence of interest in an age-related manner and/or to liver tissue, a vector is constructed in which the nucleic acid sequences of the invention (e.g., AE5' alone, AE3' alone, or a combination of AE5' and AE3') are operably linked to a promoter sequence and to a nucleotide sequence of interest. In one embodiment, the nucleotide sequence of interest is the coding region of the hFIX gene (Example 1). In another embodiment the nucleotide sequence of interest is the coding region of the protein C gene (Example 7).

The invention is not limited to coding sequences of the hFIX gene or protein C gene. Rather, any nucleotide acid sequence whose expression is desired to be regulated by sequences provided herein are contemplated to be within the scope of this invention. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes which encode a protein [e.g., reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factor genes, activator protein 1 gene, activator protein 2 gene, Sp1 gene, etc.]. In one preferred embodiment, the structural gene is the human α1-antitrypsin gene (Figure 10) (SEQ ID NO:42) which encodes a plasma proteinase inhibitor used for treating emphysema. In another preferred embodiment, the structural gene is one encoding the human antithrombin III (Figure 11) (SEQ ID NO:43) which is a plasma proteinase inhibitor for activated blood coagulation factors and whose activity is increased by heparin. In yet another preferred embodiment, the structural gene is the gene encoding the PEA-3 protein (Figure 9) (SEQ ID NO:47) and/or its related protein, which has been shown to bind specifically to homologs of the PEA-3 nucleotide sequence (SEQ ID NO:2) disclosed herein.

5

10

15

20

25

30

The invention is not limited to using a single nucleotide sequence of interest in operable combination with the sequences of the invention. Rather, a plurality (i.e., more than one) of nucleotide sequences of interest may be ligated in tandem such that their expression is regulated by the regulatory sequences of the invention. A plurality of coding sequences may be desirable, for example, where it is useful to express a transcription product of more than one gene to permit interaction of these transcriptional products. Alternatively, a plurality of coding sequences may be desirable where one of the gene sequences is a reporter gene sequence. For example, it may be advantageous to place a coding sequence of a reporter gene in tandem with the coding sequence of a gene of interest such that expression of the coding region of both the reporter gene and the gene of interest is regulated by the sequences of the invention. Expression of the reporter gene usually correlates with expression of the gene of interest. Examples of reporter gene sequences include the sequences encoding the enzymes β-galactosidase and luciferase. Fusion genes may also be desirable to facilitate purification of the expressed protein. For example, the heterologous sequence which encodes protein A allows purification of the fusion protein on immobilized immunoglobulin. Other affinity traps are well known in the art and can be utilized to advantage in purifying the expressed fusion protein. For example, pGEX vectors (Promega, Madison WI) may be used to express the polypeptides of interest as a fusion protein with

glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Other fusion polypeptides useful in the purification of proteins of interest are commercially available, including histidine tails (which bind to Ni<sup>2+</sup>), biotin (which binds to streptavidin), and maltose-binding protein (MBP) (which binds to amylose). Proteins made in such systems may be designed to include heparin, thrombin or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released at will from the heterologous polypeptide moiety to which it is fused.

5

10

15

20

25

30

One of skill in the art would understand that where a plurality of nucleotide sequences of interest is operably linked to sequences of the present invention, the nucleotide sequences of interest may be either contiguous or separated by intervening polynucleotide sequences, so long as the nucleic acid sequences of interest are operably linked to the promoter sequence, and so long as the sequences of the invention are operably linked to the promoter sequence.

While specific preferred embodiments used herein disclose the use of the hFIX promoter and the CMV promoter, it is not intended that the invention be limited to the type or source of the promoter sequence which is operably linked to the sequences of the invention. Any promoter whose activity is desired to be regulated by the sequences provided herein is contemplated to be within the scope of the invention. Exemplary promoters include the tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter (can be isolated from vector plasmid pRc/RSV from Invitrogen), retrovirus LTR promoter (can be isolated from vector plasmid pLXSN from Clontech) SV40 promoter (located between positions +3530 to +3192 in vector plasmid pCR3 from Invitrogen), PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>2</sub> promoter having the 23-bp sequence (SEQ ID NO:44) 5'-TAATACGACTCACTATAGGGCGA-3', T

gromoter having the 23-bp sequence (SEQ ID NO:45) 5'-TTATTAACCCTCACTAAAGGGAAG -3', SP6 promoter having the 23-bp sequence (SEQ ID NO:46) 5'-ATTTAGGTGACACTATAGAATAC -3', and K11 promoter. The T<sub>2</sub> promoter, T

gromoter, SP6 promoter and K11 promoter have been described in U.S Patent No.

Nor is the invention intended to be limited to the use of a single promoter. For example, chimeric promoters (i.e., two or more promoters which are derived from at least

5,591,601, the entire contents of which are incorporated by reference.

one gene) are expressly contemplated to be within the scope of the invention. Such chimeric promoters may be desirable where, for example, chimeric promoters result in increased levels of expression of an operably linked downstream coding sequence. Chimeric promoters are known in the art and include, for example, the double *Tet* promoter [Kistner et al. (1996) Proc. Natl. Acad. Sci. USA 93:10933-10938], and the U1 snRNA promoter-CMV promoter/enhancer [Bartlett et al. (1996) Proc. Natl. Acad. Sci. USA 93:8852-8857].

5

10

15

20

25

30

Expression vectors in which expression of a nucleic acid sequence of interest is regulated by sequences of the invention may be constructed using the teachings of the present invention in conjunction with techniques well known in the art. [Sambrook et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY; Ausubel et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY]. Briefly, the nucleic acid sequence of interest is placed in operable combination with a promoter sequence and sequences of the invention in the presence of transcription and translation regulatory sequences, including initiation signals such as a start codon (i.e., ATG), enhancers, and transcription termination signals. The ATG initiation codon must be in the correct reading frame to ensure translation of the entire heterologous nucleotide sequence. Transcription termination signals are placed downstream of the heterologous nucleic acid sequence and include polyadenylation sequences which are exemplified by, but not limited to, SV40 poly-A sequence, hINV poly-A sequence, or bovine growth hormone poly-A sequence, etc.

Other regulatory sequences which may affect RNA stability as well as enhancers (i.e., a sequence which when activated results in an increase in the basal rate of transcription of a gene) and silencers (i.e., a sequence involved in reducing expression of a gene) may also be included. These regulatory sequences may be relatively position-insensitive, i.e., the regulatory element will function correctly even if positioned differently in relation to the heterologous nucleotide sequence in the construct as compared to its position in relation to the corresponding heterologous nucleotide sequence in the genome. For example, an enhancer may be located at different distances from the promoter sequence, in a different orientation, and/or in a different linear order. Thus, an enhancer that is located 3' to a promoter sequence in germline configuration might be located 5' to the promoter sequence in the construct.

#### 2. Host Cells

5

10

15

20

25

30

Host cells are transformed with expression vectors which contain the sequences of the invention in operable combination with a nucleic acid sequence of interest using methods known in the art. The term "transformation" as used herein refers to the introduction of a transgene into a cell. The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations.

The term "transgene" as used herein refers to any nucleic acid sequence which is introduced into the genome of a cell by experimental manipulations. A transgene may be an "endogenous DNA sequence," or a "heterologous DNA sequence." The term "endogenous DNA sequence" refers to a nucleotide sequence which is naturally found in the cell into which it is introduced so long as it does not contain some modification (e.g., a point mutation, the presence of a selectable marker gene, etc.) relative to the naturally-occurring sequence. The terms "heterologous DNA sequence" and "foreign DNA sequence" refer to a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Heterologous DNA also includes an endogenous DNA sequence which contains some modification (e.g., a point mutation, the presence of a selectable marker gene, etc.) relative to the naturally-occurring gene. Generally, although not necessarily, heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed. Examples of heterologous DNA include reporter genes, transcriptional and translational regulatory sequences, selectable marker proteins (e.g., proteins which confer drug resistance), etc.

Transformation may be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, retroviral infection, biolistics (*i.e.*, particle bombardment) and the like.

Transformation of a cell may be stable or transient. The term "transient transformation" or "transiently transformed" refers to the introduction of one or more transgenes into a cell in the absence of integration of the transgene into the host cell's genome. Transient transformation may be detected by, for example, enzyme-linked immunosorbent assay (ELISA) which detects the presence of a polypeptide encoded by one

5

10

15

20

25

30

or more of the transgenes. Alternatively, transient transformation may be detected by detecting the activity of the protein encoded by the transgene. For example, the activity of β-glucuronidase (GUS) which is encoded by the uid A gene may be detected using either a histochemical assay of GUS enzyme activity by staining with X-gluc which gives a blue precipitate in the presence of the GUS enzyme, or a chemiluminescent assay using the GUS-Light kit (Tropix). The term "transient transformant" refers to a cell which has transiently incorporated one or more transgenes. In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more transgenes into the genome of a cell. Stable transformation of a cell may be detected by Southern blot hybridization of genomic DNA of the cell with nucleic acid sequences which are capable of binding to one or more of the transgenes. Alternatively, stable transformation of a cell may also be detected by the polymerase chain reaction (PCR) of genomic DNA of the cell to amplify transgene sequences. The term "stable transformant" refers to a cell which has stably integrated one or more transgenes into the genomic DNA. Thus, a stable transformant is distinguished from a transient transformant in that, whereas genomic DNA from the stable transformant contains one or more transgenes, genomic DNA from the transient transformant does not contain a transgene.

Suitable host cells include bacterial, yeast, plant, insect, and mammalian cells. In one embodiment the host cell is mammalian. In a preferred embodiment, the mammalian host cell is a mouse fertilized egg cell. In an alternative embodiment, the mammalian host cell is a HepG2 cell (ATCC number HB8065), a fibroblast cell (e.g., ATCC number CCL 110), a myoblast cell (e.g., Clonetics, catalog # SkMC), and an endothelial cell (e.g., human umbilical cord endothelial cells; ATCC number CRL 1730).

In one embodiment, the host cell is transformed both with an expression vector which contains the sequences of the invention in operable combination with the nucleic acid sequences of interest, as well as with an expression vector which expresses the PEA-3 protein (Example 6). Such co-transformation may be desirable, for example, where expression of the nucleotide sequence of interest is regulated by AE5' or portions or homologs thereof which contain homologs of the PEA-3 nucleotide sequence to which the PEA-3 protein binds. In one embodiment, expression of the PEA-3 protein is under the control of the LTR promoter of the Moloney murine leukemia virus (MoLV) which is capable of driving expression of operably linked genes in several cell types. Transient

expression assays are suitable for determining the relative promoter activities in expressing desirable PEA-3 protein levels.

5

10

15

20

25

30

Any number of selection systems may be used to recover transfected cells. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy I et al. (1980) Cell 22:817-23) genes which can be employed in the or aprt cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate [Wigler M et al., (1980) Proc Natl Acad Sci 77:3567-70]; npt, which confers resistance to the aminoglycosides neomycin and G-418 [Colbere-Garapin F et al., (1981) J. Mol. Biol. 150:1-14] and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine [Hartman SC and RC Mulligan (1988) Proc Natl Acad Sci 85:8047-51]. Recently, the use of a reporter gene system which expresses visible markers has gained popularity with such markers as β-glucuronidase and its substrate (GUS), luciferase and its substrate (luciferin), and β-galactosidase and its substrate (X-Gal) being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system [Rhodes CA et al. (1995) Methods Mol Biol 55:121-131].

The presence or expression of the reporter gene usually indicates the presence or expression, respectively, of the tandem heterologous nucleic acid sequence as well. However, it is preferred that the presence and expression of the desired heterologous nucleic acid sequence be confirmed. This is accomplished by procedures known in the art which include DNA-DNA or DNA-RNA hybridization or amplification using probes, or fragments of the heterologous nucleic acid sequence. For example, Fluorescent In Situ Hybridization (FISH) can be used to detect the heterologous nucleic acid sequence in cells. Several guides to FISH techniques are available, e.g., Gall et al. Meth. Enzymol. 21:470-480 (1981); Angerer et al., in "Genetic Engineering: Principles and Methods," Setlow & Hollaender, Eds. Vol. 7 pp. 43-65, Plenum Press, New York (1985). Alternatively, DNA or RNA can be isolated from cells for detection of the transgene by Southern or Northern hybridization or by amplification based assays. Nucleic acid amplification based assays involve the use of

oligonucleotides or oligomers based on the nucleotide sequence of interest in order to detect cells and tissues which contain the DNA or RNA encoding the transgene of interest. As used herein, the terms "oligonucleotides" and "oligomers" refer to a nucleic acid sequence of at least about five (5) contiguous nucleotide residues and as many as about sixty (60) nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplimer. Standard PCR methods useful in the present invention are described by Innis *et al.* (Eds.), "PCR Protocols: A Guide to Methods and Applications," Academic Press, San Diego (1990).

5

10

15

20

25

30

Yet another alternative for the detection of heterologous nucleic acid sequences includes detecting the polypeptide product of transcription of the heterologous nucleotide sequence. A variety of protocols which employ polyclonal or monoclonal antibodies specific for the protein product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A competitive binding assay may also be used. Alternatively, a two-site, monoclonal-based immunoassay which utilizes monoclonal antibodies that are reactive to two non-interfering epitopes on the protein of interest may be employed. These and other assays are described in, among other places, Hampton R et al. (1990), Serological Methods a Laboratory Manual, APS Press, St Paul MN), and Maddox DE et al. (1983), J. Exp. Med. 158:1211.

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting related sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the nucleotide sequence of interest, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides. A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like.

Host cells transformed with expression vectors containing the sequences provided herein are useful for age-related expression of recombinant proteins of interest. Host cells transformed with expression vectors containing the invention's sequences may be part of a tissue or organ of a living animal. A "living animal" as used herein refers to any multicellular animal (e.g., humans, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc.) into whose cells the sequences provided herein may be introduced. Where the host cells (e.g., fertilized egg cells) are capable of generating a multicellular organism, these cells when transformed with expression vectors containing the sequences of the invention are useful in generating transgenic animals which exhibit age-related and/or liver-specific expression of nucleotide sequences of interest.

# 3. Transgenic Animals

5

10

15

20

25

30

The present invention provides transgenic non-human animals which express a nucleotide sequence of interest in an age-related manner. These animals provide useful models for diseases (e.g., thrombosis, cardiovascular diseases, diabetes, Alzheimer's disease, cancer, osteoporosis, osteoarthritis, Parkinson's disease, dementia) which are associated with increasing age, as well a for screening candidate therapeutic agents against such diseases. These transgenic animals are also useful in studies of normal phenomena, such as ageing, gene regulation, etc. In one embodiment, the invention discloses transgenic mice which express in an age-related manner the exemplary hFIX coding sequence under the control of AE5' and/or AE3' (Example 3).

The term "age-related manner" when made in reference to the expression of a nucleotide sequence of interest is a relative term which refers to an increase over a period of time in the quantity of mRNA and/or protein encoded by the nucleotide sequence of interest when the nucleotide sequence of interest is operably linked to a promoter and to a nucleic acid sequence which has age-related regulatory activity, as compared to the quantity of mRNA and/or protein, respectively, encoded by the nucleotide sequence of interest when the nucleotide sequence of interest is operably linked to the promoter in the absence of the nucleic acid sequence which has age-related regulatory activity. Thus, the term "age-related" when made in reference to expression of a nucleotide sequence of interest by a transgenic

animal means that the transgenic animal expresses the nucleotide sequence of interest in an age-related manner.

For example, in one embodiment, the invention demonstrates that hFIX is expressed in an age-related manner in transgenic mice which harbor a transgene (-416FIXm1/1.4) (Figure 2B) which contains hFIX under the control of the hFIX promoter and the regulatory control of AE3' as compared to expression of hFIX in transgenic mice which harbor a transgene (-416FIXm1) (Figure 2A) in which hFIX is under the control of the hFIX promoter in the absence of AE3'. While transgenic mice harboring the -416FIXm1/1.4 construct showed decreasing hFIX activity levels over a period of time (e.g., from 1 to 9 months of age), this decrease was less than the decrease in hFIX activity levels which was observed in transgenic mice harboring the -416FIXm1 construct over the same period of time.

5

10

15

20

25

30

In another embodiment, the invention discloses that hFIX is expressed in an age-related manner in transgenic mice which harbor transgenes (-802FIXm1, -2231FIXm1, and -416FIXm1/AE5') (Figures 4A, C and E) each of which contains hFIX under the control of the hFIX promoter and the regulatory control of AE5' as compared to expression of hFIX in transgenic mice which harbor a transgenes (-416FIXm1 and -770FIXm1) (Figures 2A and E) in which hFIX is under the control of the hFIX promoter in the absence of AE5'.

Transgenic mice harboring each of the -802FIXm1, -2231FIXm1, and -416FIXm1/AE5' constructs showed relatively unchanged hFIX activity levels over a period of time (e.g., from 1 to 7 months of age) while transgenic mice harboring either the -416FIXm1 or -770FIXm1 construct showed decreasing hFIX activity levels over the same time period.

In an additional embodiment, the invention shows that hFIX is expressed in an agerelated manner in transgenic mice which harbor transgenes (-802FIXm1/1.4 and
-2231FIXm1/1.4) (Figures 4B and D) each of which contains hFIX under the control of the
hFIX promoter and the regulatory control of both AE3' and AE5' as compared to expression
of hFIX in transgenic mice which harbor a transgene (-770FIXm1) (Figure 2E) in which
hFIX is under the control of the hFIX promoter in the absence of both AE3' and AE5'.

Transgenic mice harboring either the -802FIXm1/1.4 or the -2231FIXm1/1.4 construct
showed increasing levels of hFIX activity over a period of time (e.g., 1 to 3 months of age)
as compared to decreasing hFIX activity levels over the same period of time in transgenic
mice harboring the -770FIXm1 construct.

The present invention also provides transgenic non-human animals which express a nucleotide sequence of interest in a liver-specific manner. These animals are useful for targeting expression of a nucleotide sequence of interest to the liver. Examples of nucleotide sequences of interest are those which encode blood coagulation factors (e.g., factor VIII, factor VII, factor X and prothrombin) whose deficiency is known to play a role in abnormal bleeding disorders. Other examples of nucleotide sequences of interest include those which encode blood coagulation regulators and/or inhibitors (e.g., protein C, antithrombin III, and tissue factor pathway inhibitor [TFPI]) whose deficiency results in thrombosis,  $\alpha$ 1-antitrypsin whose deficiency results in emhysima, and LDL-receptor whose deficiency results in hypercholestrolemia.

5

10

15

20

25

30

Yet other examples of a nucleotide sequence of interest include those encoding enzymes involved in specific metabolic defects (e.g., urea cycle enzymes, especially ornithine transcarbamylase, argininosuccinate synthase, and carbamyl phosphate synthase); receptors (e.g., LDL receptor); toxins; thymidine kinase to ablate specific cells or tissues; ion channels (e.g., chloride channel of cystic fibrosis); membrane transporters (e.g., glucose transporter); and cytoskeletal proteins (e.g., dystrophin). The nucleotide sequence of interest may be of synthetic, cDNA, or genomic origin, or a combination thereof. The nucleotide sequence of interest may be one which occurs in nature, a non-naturally occurring gene which nonetheless encodes a naturally occurring polypeptide, or a gene which encodes a recognizable mutant of such a polypeptide. It may also encode an mRNA which will be "antisense" to a DNA found or to an mRNA normally transcribed in the host cell, but which antisense RNA is not itself translatable into a protein. In one embodiment, the invention discloses transgenic mice which express in a liver-specific manner the exemplary hFIX coding sequence under the control of AE5' (Example 3).

The term "liver-specific manner" as used herein in reference to the expression of a nucleotide sequence of interest in a transgenic animal is a relative term which means that the quantity of mRNA and/or protein encoded in liver tissue by the nucleotide sequence of interest is greater than, preferably two times greater, more preferably five times greater, and most preferably ten times greater, than the quantity of mRNA and/or protein encoded by the nucleotide sequence of interest in tissues other than liver tissue of the same transgenic animal as detected by Northern blot hybridization and/or by the activity of the encoded protein as described herein. Thus, the term "liver-specific" when made in reference to expression of a

nucleotide sequence of interest by a transgenic animal means that the transgenic animal expresses the nucleotide sequence of interest in an liver-specific manner.

A first step in the generation of the transgenic animals of the invention is the introduction of a construct containing nucleic acid sequences of interest under the expression regulatory control of sequences of the invention into target cells. Several methods are available for accomplishing this, including microinjection, retroviral infection, and implantation of embryonic stem cells. These methods are discussed as follows.

# i. Microinjection Methods

10

5

Direct microinjection of expression vectors into pronuclei of fertilized eggs is the preferred, and most prevalent, technique for introducing heterologous nucleic acid sequences into the germ line. Technical aspects of the microinjection procedure and important parameters for optimizing integration of nucleic acid sequences have been previously described [Hogan *et al.*, (1986) Manipulation of the Mouse Embryo: A Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Lab.].

15

20

Once the expression vector has been injected into the fertilized egg cell, the cell is implanted into the uterus of a pseudopregnant female and allowed to develop into an animal. Of the founder transgenic animals born, 70% carry the expression vector sequence in all of their cells, including the germ cells. The remaining 30% of the transgenic animals are chimeric in somatic and germ cells because integration of the expression vector sequence occurs after one or more rounds of replication. Heterozygous and homozygous animals can then be produced by interbreeding founder transgenics. This method has been successful in producing transgenic mice, sheep, pigs, rabbits and cattle [Hammer et al., (1986) J. Animal Sci.:63:269; Hammer et al., (1985) Nature 315:680-683].

25

30

#### ii. Retroviral Methods

Retroviral infection of pre-implantation embryos with genetically engineered retroviruses may also be used to introduce transgenes into an animal cell. For example, blastomeres have been used as targets for retroviral infection [Jaenisch, (1976) Proc. Natl. Acad. Sci USA 73:1260-1264]. Transfection is typically achieved using a replication-defective retrovirus carrying the transgene [Jahner *et al.*, (1985) Proc. Natl. Acad. Sci. USA 82:6927-6931; Van der Putten *et al.*, (1985) Proc. Natl. Acad Sci USA 82:6148-6152].

Transfection is obtained, for example, by culturing eight-cell embryos, from which the zona pellucida has been removed with fibroblasts which produce the virus [Van der Putten (1985), supra; Stewart et al., (1987) EMBO J. 6:383-388]. The transfected embryos are then transferred to foster mothers for continued development. Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele [Jahner et al., (1982) Nature 298:623-628]. Yet another alternative method involves intrauterine retroviral infection of the midgestation embryos [Jahner et al. (1982), supra].

The advantages of retroviral infection methods include the ease of transfection and the insertion of a single copy of the transgene, which is flanked by the retroviral long terminal repeats (LTRs), into the chromosome. However, this method is not a preferred method because most of the founders will show mosaicism since infection occurs after cell division has begun. This necessitates outbreeding to establish homozygous and heterozygous lines suitable for analysis of gene expression. More importantly, the retroviral LTR sequences may interfere with the activity of the hINV upstream sequences in directing expression of the heterologous nucleic aid sequences.

### iii. Embryonic Stem Cell Implantation

5

10

15

20

25

30

Another method of introducing transgenes into the germ line involves using embryonic stem (ES) cells as recipients of the expression vector. ES cells are pluripotent cells directly derived from the inner cell mass of blastocysts [Doetchman et al., (1988) Dev. Biol. 127:224-227], from inner cell masses [Tokunaga et al., (1989) Jpn. J. Anim. Reprod. 35:173-178], from disaggregated morulae [Eistetter, (1989) Dev. Gro. Differ. 31:275-282] or from primordial germ cells [Matsui et al., (1992) Cell 70:841-847]. Expression vectors can be introduced into ES cells using any method which is suitable for gene transfer into cells, e.g., by transfection, cell fusion, electroporation, microinjection, DNA viruses, and RNA viruses [Johnson et al., (1989) Fetal Ther. 4 (Suppl. 1):28-39].

The advantages of using ES cells include their ability to form permanent cell lines in vitro, thus providing an unlimited source of genetic material. Additionally ES cells are the most pluripotent cultured animal cells known. For example, when ES cells are injected into an intact blastocyst cavity or under the zona pellucida, at the morula stage embryo, ES cells are capable of contributing to all somatic tissues including the germ line in the resulting chimeras.

Once the expression vector has been introduced into an ES cell, the modified ES cell is then introduced back into the embryonic environment for expression and subsequent transmission to progeny animals. The most commonly used method is the injection of several ES cells into the blastocoel cavity of intact blastocysts [Bradley et al., (1984) Nature 309:225-256]. Alternatively, a clump of ES cells may be sandwiched between two eight-cell embryos [Bradley et al., (1987) in "Teratocarcinomas and Embryonic Stem Cells: A Practical Approach," Ed. Robertson E.J. (IRL, Oxford, U.K.), pp. 113-151; Nagy et al., (1990) Development 110:815-821]. Both methods result in germ line transmission at high frequency.

10

5

Target cells which contain the heterologous nucleic acid sequences are recovered, and the presence of the heterologous nucleic acid sequence in the target cells as well as in the animal is accomplished as described *supra*.

# E. Gene Therapy

15

20

25

30

The regulatory nucleic acid sequences provided herein may be used for gene therapy applications in both non-human animals as well as in humans. For example, the regulatory nucleic acid sequences of the invention may be introduced into cells using an expression vector which encodes a polypeptide sequence of interest using a variety of means known in the art to be useful both for delivery *in vivo* and *ex vivo*, including (1) recombinant retroviral transduction, (2) recombinant adenoviral vectors, (3) targeted cationic liposomes, and (4) gene transfer using biolistics, as described in the following sections.

### 1. Recombinant Retroviral Transduction

Retroviral vectors encoding polypeptides of interest may be used for the expression of the polypeptides in any desired cell, such as primary tumor cells. The transfer of polypeptides of interest using retroviruses may be made more efficient by increasing the titer of the virus encoding the polypeptides of interest and increasing the transduction efficiency. To increase the virus titer, the retroviral construct may be designed to include a selectable marker (e.g., neo gene), and cells harboring the retroviral construct are selected by growth in the presence of a suitable selective agent (e.g., G418) followed by expansion of clones producing the highest titers of virus. To improve the transduction efficiency, retrovirus are

used in combination with liposomes or poly-L-ornithine or polylysine to enhance virus uptake.

Another way to improve gene transfer efficiency using retroviruses is to increase the targeting efficiency. Many tumor cells including glioblastomas and melanomas express excess levels of the transferrin receptor. Transferrin has been used to increase the transduction efficiency of adenovirus in combination with polylysine. Several recent reports demonstrated that replacing the SU (surface) domain of the env gene of a retrovirus can increase receptor-mediated transduction efficiency. The human transferrin gene is 2097 bp long and its insertion into the SU domain of the env gene of MLV vector may not produce a stable Env product. However, since earlier studies have suggested that the modified Env fusion protein requires the native Env for stable assembly and efficient entry, co-transfection of the transferrin-env fusion gene with the native env gene may be used to produce retrovirus particles bearing a mixture of wild type and recombinant Env. The gene transfer efficiency of the new vector may be examined by transducing tumor cells expressing high levels of transferrin receptor.

#### 2. Recombinant Adenoviral Vectors

5

10

15

20

25

30

Recombinant adenoviruses can accommodate relatively large segments of foreign DNA (~7 kb), and have the advantage of a broad host cell range and high titer virus production. Adenoviruses have been used *in vivo* in rats to efficiently deliver genes to the liver and the pancreatic islets [reviewed in Becker *et al.* (1994) In *Protein Expression in Animal Cells*, Roth *et al.* eds.] and to the central nervous system [Davidson *et al.* (1993) Nature Genet. 3:219]. Rat livers have also been efficiently transduced *ex vivo* and then re-implanted [Shaked *et al.* (1994) Transplantation 57:1508]. Thus, the present invention contemplates *ex vivo* transfection followed by transplantation of the transfected cells or organ.

The replication defective recombinant adenoviruses are preferably employed; these viruses contain a deletion of the key immediate early genes E1a and E1b. To generate and propagate recombinant viruses, a packaging cell line such as 293 cells which supply the E1a and E2a proteins *in trans* is employed. Recombinant adenoviruses are created by making use of intracellular recombination between a much larger plasmid encoding most of the viral genome and a small plasmid containing the nucleotide sequence of interest flanked by

regions of homology with the viral integration site. Standard methods may be used to construct the recombinant adenoviruses [Graham and Prevec (1991) Meth. Mol. Biol. 7:109-128; Becker et al. (1994) In Protein Expression in Animal Cells, Roth et al. eds.]. Briefly, each plasmid is co-transfected together with pJM17 (Microbix Systems, Toronto) into sub-confluent monolayers of 293 cells (ATCC CRL 1573) using calcium phosphate precipitation and a glycerol shock. Initial recombinant viral stocks are titered on monolayers of 293 cells, and isolated single plaques are obtained and tested for expression of the polypeptide of interest using ELISA. Viral stocks are amplified and titered on 293 cells, and stored in aliquots at -70°C; if necessary, stocks are concentrated by centrifugation on density gradients. To infect tumor cells with recombinant adenoviruses, freshly isolated tumor cells are mixed with adenoviral stocks in a minimal volume. Titers of stocks are typically 10<sup>5</sup>- 10<sup>8</sup>/ml. Medium is replaced after several hours and the cells are followed for expression of the recombinant adenoviral-encoded polypeptide of interest (e.g., reporter genes).

A potential drawback of using an adenoviral delivery system is that the transduced cells may retain or express small quantities of adenoviral antigens on their surface. "Second generation" adenoviral vectors which contain deletions in the E2a gene are available and are associated with less inflammation in the recipient and a longer period of expression of the gene of interest [Engelhardt *et al.* (1994) Proc. Natl. Acad. Sci. USA 91:6196]. If necessary, nucleic acid sequences encoding polypeptides of interest are inserted into second generation adenoviral vectors.

Recently, adenoassociated virus (AAV) vectors and chimeric lentivirus vectors have also been shown promise in the expression of polypeptide sequences of interest.

#### 3. Targeted Cationic Liposomes

5

10

15

20

25

30

Cationic liposomes have proven to be a safe and effective means for inducing the transient expression of DNA in target cells [Ledley (1995) Human Gene Ther. 6:1129]. Clinical trials are underway using cationic liposomes to introduce the CFTR gene into the lungs of cystic fibrosis patients [Caplen et al. (1994) Gene Ther. 1:139 and Alton et al. (1993) Nature Genet. 5:135] or to introduce, by direct intra-tumor injection, the T cell costimulator B7-1 into malignant melanoma lesions in order to induce a cell-mediated immune response [Nabel et al. (1993) Proc. Natl. Acad. Sci. USA 90:11307].

Cationic liposomes (e.g., DOTAP/DOPE) and ligand-targeted cationic liposomes may be employed for the delivery of polypeptides of interest to tumor cells. Recently, in addition to cationic liposomes, neutral liposomes have also been reproted to also be useful in targeing ligands to cells. Ligand-targeted liposomes are made by covalently attaching ligands or antibodies to the surface of the cationic liposome. For example, when glioblastoma cells are to be targeted, transferrin is used as the ligand as glioblastoma cells express high levels of the transferrin receptor on their surface. When melanoma cells are to be targeted, internalizing receptors, monoclonal antibodies directed against melanoma-specific surface antigens (e.g., mAb HMSA5) may be employed as the ligand.

10

15

20

5

Plasmid DNA encoding polypeptides of interest is formed into a complex with preformed cationic liposomes using standard methodology or alternatively the DNA is encapsulated into the liposome interior. The DNA-containing liposomes are then used to transfer the DNA to tumor cells *in vivo* by direct intra-tumor injection or *in vitro* (using freshly explanted tumor cells) followed by return of the transduced cells to the recipient (e.g., a human patient or non-human animal).

### 4. Gene Transfer Using Biolistics

Biolistics (microballistics) is a method of delivering DNA into cells by projection of DNA-coated particles into cells or tissues. DNA is coated onto the surface of gold or tungsten microparticles (~1-3 µm diameter) and these particles are accelerated to high velocity and are impacted onto the target cells. The particles burst through the cell membrane and lodge within the target cell. The cell membrane quickly reseals and the passenger DNA elutes off of the particle and is expressed. The biolistic method has been used to transfect mammalian cells [Sanford *et al.* (1993) Methods Enzymol. 217:483].

25

A hand-held biolistic apparatus (BioRad) is used to transfer DNA into tumor cells or isolated tumor fragments. This device uses compressed helium to drive a disc-shaped macroprojectile which carries on its surface microparticles (1-5 µm) of gold which have been coated with purified plasmid DNA (coprecipitated with spermine) (Williams *et al.*, *supra*). This apparatus has been used to successfully transfect primary tissues.

30

Plasmid DNA encoding the polypeptides of interest may be coated onto the surface of gold microparticles according to the manufacturer's instructions (BioRad) and the biolistic apparatus is used to transfer the DNA into freshly explanted tumor cells or directly into

exposed tumors (e.g., metastatic nodules on the surface of the liver, melanoma lesions on the skin).

Regardless of the method of delivery of the expression vector into a cell, it is preferred, though not required, that the expression vector contain a selection marker (e.g., neo gene) to facilitate selection of transfected cells. Transfected cells are selected by growth in the presence of G418 (e.g., 200 µg/ml), followed by culture in growth medium containing reduced concentrations of G418 (e.g., 100 µg/ml) and growth to confluence. Expression of the polypeptides of interest is evaluated using, for example, immunoblot analysis or flow cytometry using monoclonal antibodies which are specific for the polypeptides of interest. It is preferred, though not necessary, that expression of the polypeptides of interest in the transfected tumor cells is both constitutive and stable. Constitutive expression refers to expression in the absence of a triggering event or condition, and can be achieved by the selection of a promoter which drives expression of the nucleic acid sequence encoding the polypeptides of interest. Examples of promoters which drive constitutive expression of a structural nucleic acid sequence which is operably linked to the promoter include the SR $\alpha$  promoter, CMV promoter, and HIV promoter.

Regardless of the type of expression vector used for delivery of the nucleic acid sequences of interest into a cell, the expression vector may be introduced to the cell by direct injection into tumor and/or preneoplastic tissue, systemic (e.g., intravenous) administration, aerosol administration (e.g., for delivery to the bronchial tree and other lung tissues), injection into breast ducts (e.g., for delivery to breast tissue), and topical administration (e.g., for delivery to cervical tissue).

#### F. Reducing Expression Of Factor IX In An Animal

5

10

15

20

25

30

The regulatory sequences of the invention may also be used to reduce expression of a polypeptide sequence of interest which is encoded by a nucleic acid sequence whose transcription is under the regulatory control of the regulatory sequences provided herein. For example, the regulatory sequences of the invention may be used to reduce the rate of agerelated increase of EIX activity in an animal as a means of treating diseases (e.g., thrombosis, cardiovascular disease, etc.) which are associated with age-related increases in FIX activity. Since the inventors have discovered that the exemplary nucleic acid sequences AE5' and AE3' regulate stable and increased expression levels, respectively, of hFIX, the

increase in the level of hFIX activity over time may be reduced by inhibiting the function of AE3' which regulates increased expression of hFIX. This approach has the advantage that expression of hFIX remains under the control of AE5' thus providing hFIX activities which are stable over time and which continue to play an important role in normal blood coagulation processes.

5

10

15

20

25

**30** 

The function of AE3' in age-related expression of FIX may be inhibited by, for example, inhibiting the activity of the protein which specifically binds to AE3'. The protein(s) which bind to AE3' may be identified by using the AE3' (or the minimum portion of AE3' which has age-related regulatory activity) to screen protein libraries for specific binding to AE3' or its portion. Once the protein which binds to AE3' is identified, the function of this protein may be inhibited using antibodies which are specific for this protein. Antibodies which are specific for the protein which binds to AE3' are expected to disrupt the interaction between AE3' and this protein.

Antibodies (polyclonal and monoclonal) which are specific for the protein that binds to AE3' or portions thereof may be generated using methods known in the art. The term "antibody" refers to immunoglobulin evoked in animals by an immunogen (antigen). It is desired that the antibody demonstrates specificity to epitopes contained in the immunogen. The term "polyclonal antibody" refers to immunoglobulin produced from more than a single clone of plasma cells; in contrast "monoclonal antibody" refers to immunoglobulin produced from a single clone of plasma cells. The terms "specific binding," "specifically binding" and grammatical equivalents thereof when used in reference to the interaction of an antibody and an immunogen mean that the interaction is dependent upon the presence of a particular structure (*i.e.*, the antigenic determinant or epitope) on the immunogen; in other words the antibody is recognizing and binding to a specific immunogen structure rather than to immunogens in general. For example, if an antibody is specific for epitope "A", the presence of an immunogen containing epitope A (or free, unlabelled A) in a reaction containing labelled "A" and the antibody will reduce the amount of labelled A bound to the antibody.

Polyclonal and monoclonal antibodies which are specific to a desirable polypeptide, given the teachings herein, may readily be prepared by one of skill in the art. For example, monoclonal antibodies may be generated by immunizing an animal (e.g., mouse, rabbit, etc.) with a desired antigen and the spleen cells from the immunized animal are immortalized,

commonly by fusion with a myeloma cell. Immunization with antigen may be accomplished in the presence or absence of an adjuvant, e.g., Freund's adjuvant. Typically, for a mouse, 10 µg antigen in 50-200 µl adjuvant or aqueous solution is administered per mouse by subcutaneous, intraperitoneal or intra-muscular routes. Booster immunization may be given at intervals, e.g., 2-8 weeks. The final boost is given approximately 2-4 days prior to fusion and is generally given in aqueous form rather than in adjuvant.

5

10

15

20

25

30

Spleen cells from the immunized animals may be prepared by teasing the spleen through a sterile sieve into culture medium at room temperature, or by gently releasing the spleen cells into medium by pressure between the frosted ends of two sterile glass microscope slides. The cells are harvested by centrifugation (400 x g for 5 min.), washed and counted. Spleen cells are fused with myeloma cells to generate hybridoma cell lines. Several mouse myeloma cell lines which have been selected for sensitivity to hypoxanthineaminopterin-thymidine (HAT) are commercially available and may be grown in, for example, Dulbecco's modified Eagle's medium (DMEM) (Gibco BRL) containing 10-15% fetal calf serum. Fusion of myeloma cells and spleen cells may be accomplished using polyethylene glycol (PEG) or by electrofusion using protocols which are routine in the art. Fused cells are distributed into 96-well plates followed by selection of fused cells by culture for 1-2 weeks in 0.1 ml DMEM containing 10-15% fetal calf serum and HAT. The supernatants are screened for antibody production using methods well known in the art. Hybridoma clones from wells containing cells which produce antibody are obtained, e.g., by limiting dilution. Cloned hybridoma cells (4-5 x 106) are implanted intraperitoneally in recipient mice, preferably of a BALB/c genetic background. Sera and ascites fluids are collected from mice after 10-14 days.

The invention also contemplates humanized antibodies which may be generated using methods known in the art, such as those described in U.S. Patent Numbers 5,545,806; 5,569,825 and 5,625,126, the entire contents os which are incorporated by reference. Such methods include, for example, generation of transgenic non-human animals which contain human immunoglobulin chain genes and which are capable of expressing these genes to produce a repertoire of antibodies of various isotypes encoded by the human immunoglobulin genes.

5

10

15

20

25

30

PCT/US00/15728

Alternatively, the function of AE3' in age-related expression of FIX may be inhibited by, for example, inhibiting the activity of AE3' using antisense sequences which are directed to AE3'. The term "antisense" as used herein refers to a deoxyribonucleotide sequence whose sequence of deoxyribonucleotide residues is in reverse 5' to 3' orientation in relation to the sequence of deoxyribonucleotide residues in a strand of a DNA duplex. AE3' antisense sequences may be used to turn off genes under the expression regulation of AE3' by transfecting a cell or tissue with expression vectors which express high levels of a desired AE3' antisense oligomer (e.g., 15-20 nucleotides) or larger fragment. Such constructs can flood cells with antisense sequences which inhibit expression of FIX. Antisense sequences can be designed from various locations along the AE3' sequence. Animals (e.g., mice) treated with vectors expressing AE3' antisense sequences are monitored for changes in the age-related symptoms associated with FIX expression. The alleviation or treatment of one or more of these symptoms in animal by an antisense sequence suggests that the antisense sequence may be useful in the treatment and/or prevention of age-related FIX expression in humans.

#### **EXPERIMENTAL**

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof. Unless otherwise mentioned, all reference to nucleotide numbers with respect to the factor IX nucleotide sequence, refers to the nucleotide numbers of the hFIX gene sequence shown in Figure 8.

#### **EXAMPLE 1**

# Construction Of A Series of Twelve Exemplary Human Factor IX (hFIX) Minigene Expression Vectors

To explore the molecular mechanisms underlying age-related regulation of Factor IX, a series of twelve hFIX minigene expression vectors were constructed. These vectors were first analyzed *in vitro* in HepG2 cells, a human hepatoma cell line (see Example 2, *infra*). Transgenic mice harboring the hFIX minigene vectors were generated and longitudinal analyses of hFIX expression for the entire life spans of founders and successive generations of transgenic mice were carried out (See Example 3, *infra*).

The twelve exemplary minigenes contained sequences derived entirely from the hFIX gene sequence, including (a) promoter sequences of various lengths spanning up to nucleotide (nt) -2231 in the 5' flanking region, (b) the coding region containing a first intron in which the first intron's middle portion is truncated. i.e., nt +1098 through nt +5882 of Figure 8, and (c) either the complete 3' UTR sequence or the 3' UTR sequence in which the middle portion was deleted. Figure 1 shows the structure of eleven out of the twelve human FIX minigene expression constructs. The name of each construct is shown at left. The structure is depicted with the promoter-containing regions (solid thick line on left) with the 5' terminal nucleotide number. Transcribed hFIX regions (open rectangles connected with thin lines representing the shortened first intron) are followed by 3' flanking sequence regions (solid thick line at right). Arrow: transcription start site; asterisk: translation stop codon; pA: polyadenylation; sl: potential stem-loop forming dinucleotide repeats.

5

10

15

20

25

30

Construction of hFIX minigene expression vectors was carried out using -416FIXm1 as the starting construct (Kurachi et al. (1995) J. Biol. Chem. 270:5276-5281). The nucleotide (nt) numbering system used in this study was based on the complete hFIX gene sequence previously reported (Yoshitake et al. (1985) Biochem. 24:3736-3750). Minigene -416FIXm1/1.4 was constructed from -416FIXm1 by inserting the middle portion of the 3' UTR (1.2 kb in size) which was generated by PCR using the following primer set with BamH I linkers: 5' primer, TAACAGGATCCGGCCTCTCACTAACTAATCAC (nt +31418 through +31438) (SEQ ID NO:14) and 3' primer, CAACTGGATCCAAGATTCAAGATAGAAGGAT (nt +32690 through +32671) (SEQ ID NO:15), and human genomic DNA as an amplification template. The PCR product was digested with BamH I, and the generated fragment was inserted into the 3' UTR BamH I site of -416FIXm1, thus producing -416FIXm1/1.4 which contained the entire 3' UTR. -416FIX m1/0.7 was constructed by inserting the PCR-amplified 700 bp fragment with BamH I linker, containing the 3' contiguous sequence to nt +32117. The primers used were, 5' primer: same as that for -416FIXm1/1.4, 3'primer: GGACAGGATCCCC CAAACTTTTCAGGCAC (nt +32117 through +32097) (SEQ ID NO:16). Minigenes -590FIXm1, -679FIXm1, -770FIXm1, -802FIXm1 and -2231FIXm1 were produced by replacing the 5' end 433 bp sequence of -416FIXm1 released by Sph I/Nhe I digestion with 607, 696, 787, 819 and 2248 bp fragments containing the 5' end hFIX region extended up to nt -590, -679, -770, -802 and -2231, respectively. These latter fragments were generated by

Sph I/Nhe I digestion of the PCR product obtained with 5' primers:

CAAGCATGCATCTAGTGTTAGTGGAAGAG (nt -590 through -571) (SEQ ID NO:17), CAAGCATGCAAATATTAACTCAAAATGGA (nt -679 through -660) (SEQ ID NO:18), CAAGCATGCTGTTGTTTTTGTTTTAAC (nt -770 through -752) (SEQ ID NO:19), CAAGCATGCAGCCATTCAGTCGAGGAAGG (nt -802 through -783) (SEQ ID NO:20), CAAGCATGCGATCCCTTCCTTATACCT (nt -2231 through -2214) (SEQ ID NO:21) with Sph I linker and the common 3' primer TAAGCTTAACCTTTGCTAGCAGATTGT (nt +30 through +10) (SEQ ID NO:22) and human genomic DNA as the amplification template. Minigene -802FIXm1/0.7 (whose structure is not shown in Figure 1) contains the 3' UTR region through nt 32,140, which is then connected to nt 32,690 through its downstream poly (A) signal sequence that is common to each of the other eleven constructs.

-416FIXm1/AE5' depicts a construct with the AE5' region moved to the 3'-end position and shown as an open box at right. -416FIX m1/AE5' was constructed by inserting the Kpn I fragment generated by PCR (nt -802 through nt -417) into the -416FIXm1 vector (the Kpn I site is outside of the FIX gene, Figure1). The 5' and 3' primers used for PCR were CTTGGTACCAGCCATTCAGTCGAGGAAGG (nt -802 through -783) (SEQ ID NO:23) and CTTGGTACCATATGAATCCTTTCATAGAT, (nt -417 through -436) (SEQ ID NO:24) respectively. All constructs were sequenced through PCR amplified regions as well as fragment ligation sites to confirm the correct sequences and orientations.

20

5 .

10

15

#### **EXAMPLE 2**

# Transient Expression of Eleven hFIX Minigene Expression Vectors In Vitro In Human Hepatoma HepG2 Cell Line

30

25

Transient *in vitro* expression activities of hFIX minigene constructs were assayed using HepG2 cells and hFIX specific enzyme linked immunosorbent assay (ELISA) as previously described (Kurachi *et al.* (1995) J. Biol. Chem. 270:5276-5281) with some modifications. Cell transfection was carried out by the calcium phosphate-DNA conjugate method or later, using FuGene 6 (Boehringer Manheim). The latter, improved transfection method consistently increased transfection efficiency to >20% (Kurachi *et al.* (1998) Biochemica 3:43-44), and all earlier assays were reexamined using FuGene 6. Four to five independent assays of factor IX activity were carried out and the averages were shown with

standard errors. With FuGene 6 transfection, the control minigene -416FIXm1 typically produced hFIX at a level of ~50 ng/10<sup>6</sup> cells/48 hr.

5

10

15

20

25

30

Figure 1 shows the relative *in vitro* transient expression activities of the human FIX minigene expression constructs (transient expression activity of minigene -802FIXm1/0.7 which is not shown in Figure 1 was 81.5% of the activity of minigene -416FIXm1). Transient expression activities relative to the activity of -416FIXm1 (~50 ng/10<sup>6</sup> cells/48 hour, and defined as 100% activity) are shown on the right side with standard deviations (from 4-5 independent assays). Activities were normalized to the size of minigenes used.

The relative transient expression activities shown in Figure 1 show that all constructs showed comparable high transient hFIX expression in HepG2 cells (~50 ng/10<sup>6</sup> cells/48 hours). However, all the constructs containing the complete 3' UTR, including a 102 base pair (bp) stretch of inverted AT, GT and GC dinucleotide repeats [Yoshitake et al. (1985) Biochem. 24:3736-3750], reproducibly showed expression activity levels which were 25-30% lower than corresponding minigenes without the repeat sequences. Dinucleotide repeats similar to those seen in the hFIX 3' UTR, which can form stable stem-loop (sl) structures in mRNA, have been implicated in controlling mRNA stability in mammals as well as yeast and plants, thus providing an important layer of protein biosynthesis regulation [Ross (1995) Microbiol. Rev. 59:423-450]. Together, these results suggest a similar negative regulatory activity for this structure of the hFIX gene in the HepG2 assay system on expression of the hFIX gene. As described below (e.g., Example 3), however, the 3' UTR structure of the hFIX gene containing the dinucleotide repeat region showed unexpected functions in vivo which are critical for advancing age-related regulation of the hFIX gene.

Another important and surprising finding with the HepG2 cell assay system is that expression by these hFIX minigenes (which contained sequences which are positioned upstream and downstream of the hFIX gene, and which are derived from the homologous hFIX gene instead of from heterologous reporter genes) does not show any down-regulation in the presence of the 5' upstream region (nt -802 up through nt -1900) [Salier *et al.* (1990) J. Biol. Chem. 265:7062-7068] (Figure 1). In contrast, when a CAT reporter gene was used, negative regulatory elements were identified in this region [Salier *et al.* (1990) J. Biol. Chem. 265:7062-7068].

#### **EXAMPLE 3**

### Generation And Analysis Of Transgenic Mice Harboring hFIX Minigene Expression Vectors

Transgenic animals were constructed using the expression plasmids described above in Example 2 according to standard methods [Hogan et al. (1994) in "Manipulating the Mouse Embryo, a Laboratory Manual" (Cold Spring Harbor Press, New York, 2nd Edition). All animal experiments were carried out in accordance with the institutional guidelines of the University of Michigan (OPRR No. A3114-01).

Briefly, Factor IX minigene expression plasmids were double-digested with Sph I/Kpn I and the factor IX minigene-containing fragments released were isolated by 0.8% agarose gel electrophoresis, followed by purification with SpinBind DNA extraction units (FMC). Fertilized eggs of C57B/6 X SJL mice were microinjected with the DNA (1-2 ng/egg), and implanted into foster mother animals (CD-1).

### A. Multiplex PCR Analysis

5

10

15

20

25

30 ·

Offspring produced were screened for founder animals with high transgene copy numbers (5 -10 copies per genome) using quantitative multiplex PCR analyses of tail tissue DNA samples. Two pairs of primers were used, one specific to the hFIX transgenes and the other specific to mouse  $\beta$ -globin gene (endogenous control); 5' primer:

CTGTGGGAACACACAGATTTTGG (nt +6172 through +6195) (SEQ ID NO:25) and 3' primer: GGAATAATTCGAATCACAT (nt +30885 through +30867) (SEQ ID NO:26), and 5' primer: CCAATCTGCTCACACAGGAT (nt +2590 through +2609) (SEQ ID NO:27) and 3' primer: CCTTGAGGCTGTCCAAGTGA (nt +3083 through +3064) (SEQ ID NO:28), respectively. These primers were designed to amplify a unique 966 bp fragment from the hFIX transgenes and a 494 bp fragment from the mouse  $\beta$ -globin gene, respectively. PCR was initiated with 3 min incubation at 94° C, followed by 25 cycles of 94° C for 30 sec, 65° C annealing for 1 min and 72° C extension for 2 min.

Founders were back-crossed with non-transgenic mice (C57B/6 X SJL) to generate F1 progeny animals. Homozygous F2 animals were generated by crossing among heterozygous F1 littermates and the following generations were similarly generated. Zygosity status of animals was determined by quantitative multiplex PCR analysis as described above.

Minimally, three founder lines for each minigene construct were subjected to longitudinal analysis for their entire life spans up to two years.

5

10

15

20

25

30

Figure 3B shows the results of quantitative multiplex PCR analysis to determine the relative transgene levels in tail and liver tissues. Genomic DNA was extracted from snipped tail tissue of a transgenic -416FIXm1 animal (PA112) at 3 weeks and at 19 months of age. Liver DNA was extracted from the same animal (PA112) sacrificed at 19 months of age and a -416FIXm1 animal (PA412) sacrificed at 1 month of age. Positions of hFIX specific fragment (966 bp) and mouse  $\beta$ -globin specific fragment (494 bp, internal copy number control) are shown on the right. Lane 1: kb size ladder; lane 2: fragment size control amplified from -416FIXm1 plasmid; lane 3: non-transgenic mouse tail DNA as template; lane 4: tail DNA of PA112 at 3 weeks of age; lane 5: tail DNA of PA112 at 19 months of age; lane 6: liver DNA of PA412 at 1 month of age; lane 7: liver DNA of PA112 at 19 month of age. The relative transgene copy numbers for the 1 month-old versus the 19 month-old animals, normalized to the endogenous mouse  $\beta$ -globin gene, were 1.0-1.1 for both tail as well as liver genomic DNA preparations, showing no sign of loss of the hFIX transgene in the genome with age (Multi Analyst program from BioRad used for quantitation and calculation of ratios).

### B. Immunoassay of hFIX Levels In Transgenic Mice

Circulatory hFIX levels were monitored during longitudinal analyses of transgenic mice from the representative founder lines carrying various hFIX minigene transgenes. At various ages, starting at one month of age, transgenic mice were individually subjected to blood sample collection (aliquot of ~100 µl) via tail-tip snipping, and the obtained serum was routinely used to quantify hFIX levels in the circulation using duplicated hFIX-specific ELISA for each age point. Pooled human plasma (George King Bio-Medical) was used to prepare a hFIX standard curve for each assay. In order to minimize experimental fluctuations from assay to assay in the longitudinal analysis, overlapped serum samples from the previous assay group were included in each assay. To ensure reproducibility, three to six independent founder lines were generated for each minigene construct, and animals from at least three representative lines were subjected to longitudinal analyses for their entire life spans. The duplicated ELISA values varied less than 11% from the averages. The results are shown in Figures 2 and 4.

In all panels in Figures 2 and 4, labeling of animals is based on the tag numbers plus additional information. The first letters of the label F or P represent founder or progeny, respectively. Information on progeny generation (F1 or F2) and sex are in parenthesis (m: male; f: female), followed by status (+: alive in good health; d: died; s: sacrificed for various examinations; mo: age of death or sacrifice). To avoid overcrowding of the panels, the results from representative animals are shown for each minigene construct. Importantly, age-regulation patterns were remarkably similar among all animals for each specific construct and different founder lines. Panels A-E of Figure 2 show representative founder line animals with -416FIXm1 (A); -416FIXm1/1.4 (B), -590 FIXm1 (C), -679FIXm1 (D) and -770FIXm1 (E). Panels A-D of Figure 4 show representative founder line animals with -802FIXm1, -802FIXm1/1.4, -2231 FIXm1 and -2231FIXm1/1.4, respectively. Panel E shows representative founder line animals with -416FIXm1/AE5'.

5

10

15

20

25

30

Figure 2 shows that at one month of age, the mice carrying the -416FIXm1 minigene produced hFIX at varying levels, from as high as that of natural hFIX gene expression (~4 μg/ml) to much lower levels (~50 ng/ml) (Figure 2A). Such variations are primarily due to the transgene positional effects in the genome. Circulatory hFIX levels of animals from the representative founder lines carrying the minigene, however, declined drastically through puberty and during the subsequent two to three month period to much lower levels, which then remained stable for the remaining life span. This rapid age-dependent characteristic decline in the circulatory hFIX level was observed in all animals analyzed (n=69), regardless of founder line, differences in initial hFIX level at pre-pubertal age (one month) due to transgene positional effects, generation (founders and F1 or F2 progeny), sex, or zygosity status (homozygous / heterozygous) of the transgenes.

### C. Northern Blot Analysis of hFIX mRNA in Transgenic Mice

Northern blot analyses of the liver RNA samples from animals (15 µg per lane) were carried out as previously described [Kurachi et al. (1995) supra] using the <sup>32</sup>P-labeled Ssp I/BamH I fragment (the 3' half of the hFIX coding region of the cDNA) as a probe, and employing stringent washing conditions. Under these conditions, the probe preferentially hybridized strongly with hFIX minigene mRNA bands (~1.7 kb) with little cross-hybridization with the mouse FIX mRNA bands (3.2 kb and 2.2 kb) [Yao et al. (1994) Gene Therapy 1:99-107]. To confirm the presence of equivalent amounts of RNA in each lane, the filters

previously hybridized with hFIX probe were stripped of probe and re-probed with the RNR18 cDNA (ribosomal RNA 18S). After completion of longitudinal analyses of animals from key founder lines for their entire life spans, the representative lines were subjected to embryo-freezing for preservation.

5

10

15

20

25

30

The results of Northern analysis of human FIX mRNA and transgene DNA levels in the livers of animals carrying -416FIXm1 are shown in Figure 3A. hFIX mRNA levels in the liver of young (PA412: F1/f, 1 month of age) and old (PA112; F1/f, 19 months of age) transgenic animals were analyzed by Northern blot analysis of total liver RNA. PA412 and PA112 animals were from the same litter produced by the founder FA661, and expressed 1252 and 1675 ng/ml circulatory hFIX at one month of age, respectively. PA112 was expressing 63.8 ng/ml serum hFIX at the time of sacrifice. Lane 1: non-transgenic mouse liver RNA; lane 2: transgenic PA412 liver RNA; lane 3: transgenic PA112 liver RNA. FIX and 18S on the left or right sides indicate the band position of hFIXm1 mRNA (~1.7 kb) and RNR18 (1.9 kb, ribosomal RNA), respectively.

Figure 3A shows that the decline in blood hFIX level observed in Figure 2 was correlated with a similar decline in the steady-state liver hFIX mRNA, which was not due to a loss of the hFIX transgene with age (Figure 3B). This was further supported by the fact that when 4-5 month old mice with much decreased hFIX levels had progeny, their pups depicted pre-pubertal high hFIX expression levels equivalent to those of their parents at the same time point (one-month of age).

Minigene vector -416FIXm1/1.4 is identical to -416FIXm1 except that -416FIXm1/1.4 contains the complete 3'UTR, including the dinucleotide repeat structure (102 bp in length) in its middle region [Yoshitake *et al.* (1985) Biochem. 24:3736-3750] (Figure 1). Transgenic mice with -416FIXm1/1.4 (n=48) (Figure 2B) showed pre-pubertal high and subsequent age-dependent decline in hFIX levels similar to those of -416FIXm1 (Figure 2A), although the decline was less steep and expression levels were stabilized at significantly higher levels than those observed for -416FIXm1 (Figure 2B).

These results indicate that, while the 102-bp sequence containing the dinucleotide repeat structure of hFIX 3' UTR reduces the age-related decline in expression of hFIX, the presence of the complete 3' UTR containing the extensive dinucleotide repeat structure nonetheless does not completely rescue hFIX expression from the age-decline observed in all

of these animals, regardless of founder line, initial pre-pubertal hFIX level, generation, sex, or zygosity status of the transgenes.

All animals carrying minigenes -590FIXm1 and -679FIXm1 (a total of 25 and 26 animals subjected to longitudinal analysis, respectively) also showed an age-associated rapid decline in hFIX expression similar to that seen in animals carrying -416FIXm1 (Figure 2, C and D). Furthermore, hFIX expression levels in three independent founder animals generated to date carrying -770 FIXm1 also rapidly decreased over the puberty period in a similar pattern as the above minigenes (Figure 2E). These observations indicated that minigenes with the promoter region up to nt -770 contain the basic structural elements necessary for hFIX expression, but lack a structural element(s) which functions in age-associated stability of hFIX gene expression.

5

10

15

20

25

30

In contrast, striking and unexpected differences in hFIX expression patterns were observed with animals carrying the minigene -802FIXm1 (Figure 4A) as comp[ared to those carrying the minigene -416FIXm1 (Figure 2A). -802FIXm1 is composed of a vector frame identical to -416FIXm1, except that the 5' end flanking sequence included was extended to nt -802 (Figure 1).

All animals with -802FIXm1, -2231FIXm1 and -416FIXm1/AE5' (panels A, C, E) exhibited stable expression throughout their life spans. Animals with -802FIXm1/1.4 and -2231FIXm1/1.4 (Figure 4 B, D) exhibited age-associated increases in hFIX expression levels. All animals maintained or increased stable circulatory hFIX levels regardless of founder line, initial expression levels at one month of age, sex, generation or zygosity status. Mice which died at much younger ages than their normal life expectancies are marked with d. The above results show that all animals from three independent founder lines obtained with -802FIXm1 (Figure 4A) showed characteristic differences in hFIX expression pattern from animals with -416FIXm1 (Figure 2A) and -416FIXm1/1.4 (Figure 2B).

The -802FIXm1 transgenic animals (n = 62) subjected to longitudinal analysis invariably showed age-stable plasma hFIX levels for their entire life spans, mostly up to 20-24 months of age. Age-stable circulatory hFIX levels were correlated with age-stable mRNA levels (Figure 5). These observations with -802FIXm1 were further supported by age-stable hFIX expression by mice carrying -2231FIXm1 (Figure 4C). Together, these results suggest that the structural element which is responsible for age-stable expression of the hFIX gene resides in the small region spanning nt -770 through -802. We designated

this small region "age-regulatory element in the 5' end" (AE5'). This region contains a transcription factor PEA-3 nucleotide sequence (GAGGAAG: nt -784 through -790), which completely matches the consensus motif (C/G)AGGA(A/T)G [Martin et al. (1988) Proc. Natl. Acad. Sci. 85:5839-5843; Xin et al. (1992) Genes & Develop. 6:481-496; Chotteau-Lelievre et al. (1997) Oncogene 15:937-952; Gutman and Wasylyk (1990) EMBO J. 9:2241-2246]. The function of AE5' nucleotide sequence is position-independent as shown by age-stable hFIX expression by animals containing -416FIXm1/AE5', in which AE5' was moved to the 3' end outside of the hFIX minigene (Figure 4E).

5

10

15

20

25

30

Since transgenes of -416FIXm1, -590FIXm1, -679FIXm1 and -770FIXm1 differ from the minigenes -802FIXm1 and -2231FIXm1 only by their promoters, the hFIX mRNA produced from all of these minigenes (an intron spliced form of FIXm1 RNA) was expected to produce identical hFIX protein. Thus, it was hypothesized that the age-dependent decline in the circulatory hFIX level observed in animals with -416FIXm1, -590FIXm1, -679FIXm1 and -770FIXm1, but not with -802FIXm1 and -2231 FIXm1, must be due to an age-dependent decline in the transcriptional activity of the transgenes. This agrees with the facts that no significant changes with age in hFIX mRNA levels in the liver were observed for animals carrying -802FIXm1 (Figure 5, lanes 2 and 3), while advancing age-dependent declines in steady-state mRNA level were observed for -416FIXm1 (Figure 3A, lanes 2 and 3).

To further determine whether the age-dependent decline in the circulatory hFIX levels was due to an age-dependent decline in transcriptional activity of the transgenes, the effects of age on hFIX clearance from the circulation were tested as follows. Aliquots of plasmaderived hFIX preparation (5 μg/0.1 ml of PBS) were injected via tail vein into normal animals at 2, 9-10 and 19-23 months of age (n=3 per age group), which have the same genetic background as the transgenic mice (C57B/6 X SJL). The hFIX level in circulation was monitored by ELISA of collected blood samples (~50 μl aliquot) at 10 min, 2, 6, 12, 18, 24, 30, 36 and 48 hrs after protein injection. As expected, all animals of different age groups showed a typical bi-phasic clearance kinetics (two compartment distribution and clearance) with an initial rapid clearance phase (α-phase), followed by a slower clearance phase (β-phase). The results are shown in Table 1.

Table 1
Clearance Time of Human Factor IX in Mice

Age (months)	Clearance Time (T <sub>1/2</sub> of Human Factor IX)		
2	16.8 ± 0.21		
9-10	17.4 ± 0.55		
19-23	16.9 ± 0.35		

5

10

15

20

25

30

As shown in Table 1, very similar half clearance times were observed for all age groups tested. This agreed with our previous results (17.8 hours) for hFIX clearance in a different strain, BALB/c mice (2 months of age) (Yao et al. (1994) supra].

Furthermore, the results in Table 1 demonstrate that the hFIX turnover time from the circulation does not change significantly *in vivo* with increasing age, from youth (2 months), to middle age (9-10 months) to old age (19-23 months). These results further confirm that the age-dependent decline in the circulatory hFIX levels was due to an age-dependent decline in transcriptional activity of the transgenes.

It is important to note that in the *in vitro* HepG2 cell assay system, the presence or absence of AE5' in the minigenes did not make any significant difference in hFIX expression from the hFIX minigenes (Figure 1, and Example 2, *supra*). In contrast, as mentioned above in this Example, the presence or absence *in vivo* of AE5' makes a dramatic age-dependent difference in hFIX gene expression. This further demonstrates that *in vivo* longitudinal analysis is important for studying age-regulation of a gene.

Unlike the animals with -802FIXm1, mice with -802FIXm1/1.4 (which contains the complete 3'UTR) showed an advancing age-associated increase in the hFIX level in the circulation (n=48) (Figure 4 A and B). Thus, to determine whether this unexpected age-dependent increase in the circulatory hFIX level was directly correlated with an increased level of liver hFIX mRNA, Northern blot analyses of transgenic mice carrying -802FIXm1 and -802FIXm1/1.4 were conducted. The hFIX mRNA levels in the liver of 1-month (young) or 15-month (aged) mice carrying -802FIXm1 (mouse P327 or P552, respectively) and -802FIXm1/1.4 (mouse P32 and P697, respectively) are shown in Figure 5. These

animals were from the same litter produced by the founder F17549 for -802FIX m1 and F229 for -802FIXm1/1.4 (Figure 4 A and B). At the time of sacrifice, P552 and P697 were expressing 2200 and 1658 ng/ml of hFIX, respectively. The total liver RNA (15 µg from each animal was used for the Northern blot analysis performed as described in Figure 3A. Upper panel: probed with the Ssp I/BamH I fragment of hFIX cDNA; lower panel: rehybridized with RNR18 (ribosomal RNA) probe. Lane 1: non-transgenic mouse liver; lane 2: transgenic P327 liver RNA; lane 3: transgenic P552 liver RNA; lane 4: transgenic P32 liver RNA; lane 5: transgenic P697 liver RNA. PhosphorImager (Molecular Dynamics) was used for quantitation of mRNA levels (counts) and ratios of young versus old were calculated. Young and old animals carrying -802FIXm1 showed no significant differences in the mRNA level (the ratio of old over young: 0.92). In contrast, -802FIXm1/1.4 animals showed a substantial elevation in the mRNA level with older age (the ratio of old over young: 1.54).

5

10

15

20

25

30

These results (Figure 5, lanes 4 and 5) indicated the presence of another important age-regulatory nucleotide sequence, designated AE3', which is located approximately in the middle of the 3' UTR where an extensive stretch of dinucleotide repeating structures were contained. In the presence of AE5', AE3' clearly confers a crucial age-associated increase in hFIX expression. This conclusion was further supported by results obtained with -2231FIXm1 /1.4 (n=42) (Figure 4D). The unique concerted function conferred by the combination of AE5' and AE3' was again independent of founder line, initial expression levels at one month of age, sex, generation, or zygosity status of animals.

Interestingly, animals with sustained high hFIX levels in the circulation (approximately 1,500 ng/ml or higher) tended to die at a much earlier age than the expected life span (~2 years) (Figure 4 A, B, D). This happened to both males and females, but appears to be more frequent in males. Without limiting the invention to any particular mechanism, it is believed that since these transgenic mice have hFIX in addition to their own mFIX, they may be at an increased risk of lethal thrombosis compared to wild type mice which do not express the transgenes.

The above-described characterization of transgenic mice harboring hFIX transgenes demonstrates that (a) while the presence of AE5' in vitro in HepG2 cells did not affect hFIX gene expression, the presence of AE5' in vivo resulted in a dramatic age-dependent increased stability in hFIX gene expression, (b) the age-dependent decline in the circulatory hFIX level

observed in animals with -416FIXm1, -590FIXm1, -679FIXm1 and -770FIXm1 is directly correlated with the decrease in the steady-state mRNA level, which the inventors believe to be due to an age-dependent decline in the transcriptional activity of the transgenes, and (c) animals carrying -802FIXm1/1.4 shwoed a substantial elevation in the liver mRNA levels of hFIX with older age.

#### EXAMPLE 4

### Footprint And Gel Electrophoretic Mobility Shift Analysis Of The Region From Nucleotides -665 To -805 of Human Factor IX

In order to make a preliminary determination of the region within AE5' which is involved in the function of AE5', footprint analysis and gel electrophoretic mobility shift assays were performed as follows.

### A. Footprint Analysis

5

10

15

20

25

30

For footprint analysis of the region spanning from nt -665 through nt -805, the fragments used were amplified by PCR with the <sup>32</sup>P-labeled 5' primer ATGGTTAACTGACTTACGAA (nt -833 through -814) (SEQ ID NO:29) and 3' unlabeled primer GCTCCATTTTGAGTTAATATTTGTGT (nt -657 through -682) (SEQ ID NO:30). The nuclear extracts (NEs) from HepG2 human hepatoma cells and livers of young (1 month of age) and old (19 months of age) mice were prepared as previously reported [Kurachi et al. (1986) Biochemistry 33:1580-1591]. Various amounts of NEs (0, 100 and 150 μg) were incubated with the labeled fragments (30,000 CPM) for 1 hour on ice and subjected to DNase 1 digestion (0.5 unit) for 2 min at room temperature. The samples tested included those without NEs, with 100 μg and 150 μg of HepG2 cell NEs, with 100 μg and 150 μg NEs from young mice. Major and minor footprints and apparent DNase hypersensitivity sites were observed.

Footprint analysis of the region nt -665 through -805 with aged mouse liver nuclear extracts showed a major footprint (nt -784 through -802), a minor foot print (nt -721 through -728) and an interesting DNase hypersensitive region (nt -670 through -714). With nuclear extracts from one month-old animals or HepG2 cells, no such clear footprints were observed.

### B. Gel Electrophoretic Mobility Shift Assay

5

10

15

20

25

30

Gel electrophoretic mobility shift assay using mouse liver nuclear extracts from three different age groups was used. Nuclear extracts were prepared from 1, 5 or 19 month-old mouse livers (as described *supra*). Double stranded oligonucleotides containing a PEA-3 nucleotide sequence spanning from nt -797 to -776 of the hFIX gene (TTCAGTCGAGGAAGGATAGGGT) (SEQ ID NO:31) were <sup>32</sup>P-labeled at the 5' end to a specific activity of  $1.9 \times 10^9$  cpm. Aliquots of the radio-labeled oligonucleotide (20,000 cpm) were incubated with 10 µg of NEs in the presence of 1 µg of poly dI-dC in DNA binding buffer for 20 min at room temperature and subjected to polyacrylamide gel electrophoresis (Kurachi et al. (1986) *supra*). In Figure 6A, Lane 1: without NEs; lane 2: with NEs of 1 month-old mice; lane 3: with NEs of 5 month-old mice; land 4: with NEs of 19 month-old mice; lane 5: with mouse brain NEs (positive control for PEA-3, showing a slightly higher size of shifted band).

Figure 6B shows the results of the competition assay for <sup>32</sup>P-labeled double stranded oligonucleotides containing the PEA-3 nucleotide sequence. A 100-fold excess unlabeled oligonucleotide described in the preceding paragraph or mutant oligonucleotide [TTCAGTCGGTGATAGGGT (SEQ ID NO:32) with mutated sequences underlined] was incubated with 10 µg of 19 month-old mouse liver NEs for 5 min followed by addition of <sup>32</sup>P-labeled oligonucleotides as described *supra*. Lane 1: without NEs; lane 2: with NEs; lane 3: with NEs and wildtype competitor; lane 4: with NEs and mutant competitor.

In agreement with the above results of footprinting, gel electrophoretic mobility-shift (bandshift) assays showed an increase in protein binding with the nuclear extracts of aged mice (19 months of age) (Figure 6A). Bandshifts were competitively reduced with excess amounts of oligonucleotides harboring the PEA-3 motif, but not with oligonucleotides harboring a mutant PEA-3 motif sequence (Figure 6B), thus confirming the presence of the PEA-3 motif in AE5'. This is the first time that the PEA-3 protein, which is a member of the Ets family of transcription factors and which has been shown to bind to nucleotide sequences [SEQ ID NO:40; SEQ ID NO:48; and SEQ ID NO:84] that are homologous to the PEA-3 nucleotide sequence within the AE5' region [Karim et al. (1990) Genes & Develop. 4:1451-1453; Nelsen et al. (1993) Science 261:82-86; Fisher et al. (1991) Oncogene 6:2249-2254], has been implicated in such a unique role in age-stable regulation of a gene.

5

10

15

20

25

30

Without limiting the invention to any particular mechanism, the PEA-3 nucleotide sequence in the hFIX gene appears to have been generated through evolutionary drift of a L1 sequence originally recruited presumably via its retrotransposition into the 5' specific location. Modern retrotransposable L1 [Kazazian et al. (1988) Nature 332:164-166; Dombroski et al. (1993) Proc. Natl. Acad. Sci. USA 90:6513-6517; Minakami et al. (1992) Nucl. Acids Res. 20:3139-3145; Dombroskiet et al. (1994) Mol. Cell. Biol. 14:4485-4492] does not have the corresponding PEA-3 nucleotide sequence. The PEA-3 nucleotide sequence of AE5' nucleotide sequence resides within the L1-derived sequence retaining a 63-70% similarity with the ORF2 corresponding region of the modern retrotransposable L1 in the 5' to 3' orientation. Interestingly, the murine FIX gene also has the L1-derived nucleotide sequence in its 5' end region in an almost identical position as in the hFIX gene, and has multiple PEA-3 consensus nucleotide elements [Kawarura et al. in Organization of L1 Sequence in the 5' Flanking Region of Factor IX Gene [in preparation]. Age-regulation of the murine FIX gene is indeed very similar to that of the hFIX gene [Sweeney and Hoernig (1993) Am. J. Clin. Pathol. 99:687-688; Kurachi et al. (1996) Thromb. Haemost. 76:965-969], thus providing further insights into the evolutionary origin of the molecular mechanisms underlying age-associated regulation of the FIX gene.

### **EXAMPLE 5**

### Liver-Specific Expression Of The Exemplary hFIX Gene Under Control Of The hFIX Promoter

Expression of the natural FIX gene is virtually restricted to the liver [Salier et al. (1990) J. Biol. Chem. 265:7062-7068]. In order to determine whether any of the upstream and/or downstream sequences in the hFIX minigenes directed liver-specific expression of the hFIX transgene, Northern blot analysis was carried out as described *supra* (Example 3) in transgenic mice carrying -416FIXm1 and -802 FIXm1 expression vectors. Animals expressing hFIX at high level (PA412 and P580 carrying -416FIXm1 and -802FIXm1, respectively) were sacrificed at one month of age and total RNA was extracted from liver, lung, intestine, muscle, kidney, brain and heart and from untransfected HepG2 cells (negative control). The results in transgenic mice carrying -416FIXm1 and -802 FIXm1 are shown in Figure 7A and B, respectively.

In Figure 7, the positions of hFIX mRNA, RNR18 (control for RNA loading in wells), and ribosomal 28S and 18S RNA bands are shown on the left and right sides, respectively. Animals with -416FIXm/1.4 and -679FIXm1 showed tissue specific expression patterns similar to that of -416FIXm1 (A) (data not shown). Interestingly, liver expression of hFIX observed for minigenes lacking the region containing AE5' (except -770FIXm1, which remains to be tested as sufficient progeny animals become available) was high, but not as robust, as that seen with the natural gene. In addition, these minigenes expressed not only in the liver, but also in other tissues, such as kidney, lung and muscle, at various levels as high as ~20% of the liver level (Figure 7A). In clear contrast, animals with -802FIXm1 showed substantially liver-specific hFIX expression similar to that for the natural FIX gene (Figure 7B). These results suggest that the AE5' region controls liver specific expression of hFIX.

An apolipoprotein(a) transcription cotnrol region (ACR) which contains an ETS family target sequence 5'-CCCGGAAG-3' (SEQ ID NO:48) has been shown to exhibit enabled activity in vitro in liver-derived HepG2 cells. However, the ACR does not appear to be liver-specific [Yang et al. (1998), supra].

### EXAMPLE 6

### Expression Of PEA-3 Protein In HepG2 Liver Cells

20

25

30

5

10

15

Expression of the transgene FIX was observed *in vivo*, but not *in vitro* in HepG2 cells, when expression vectors containing AE5' were used (See, Examples 2 and 3, *supra*). This observation, together with the absence of a footprint in HepG2 cell NEs (See, Example 4, *supra*) suggested to the inventors that HepG2 cells' lack of expression of the FIX transgene may be a result of the cells' expression of low levels of the PEA-3 protein (and/or the PEA-3 related protein) which binds to homologs of the invention's PEA-3 nucleotide sequence. The complete human PEA-3 cDNA has not yet been cloned (the human PEA-3 cDNA sequence of GenBank accession number U18018 lacks 8 amino acids at the N-terminal region when compared to the mouse PEA-3 cDNA sequence). In order to determine the role of the PEA-3 nucleotide sequence in gene expression *in vitro*, HepG2 cells which overexpress mouse PEA-3 protein were constructed as follows.

Expression constructs containing the mouse PEA-3 cDNA sequence (GenBank Accession Number X63190; Figure 9) were constructed as follows. Using the reported mouse PEA-3 cDNA sequence three sets of PCR primers were synthesized such that the entire coding region and parts of the flanking sequences would be amplified. Reverse transcription PCR (RT-PCR) was carried out, and the amplified mouse PEA-3 cDNA sequence was inserted into an expression vector under the control of the SV40 promoter, which does not interfere with the factor IX promoter (data not shown).

The PEA-3 expression vector is used to transfect HepG2 cells using the FuGene 6 (Boehringer Manheim) since this method was shown to improve transfection efficiency (See, Example 2, supra). Transfected HepG2 cells are screened for expression of PEA-3 by Northern blot analysis and/or Western blot analysis using commercially available antibody. Transfected HepG2 cell lines which stably express PEA-3 protein are selected for further use, e.g., to analyze the underlying mechanism of PEA-3 action.

15 EXAMPLE 7

5

10

20

25

30

# In vitro And In Vivo Expression Of Exemplary Human Protein C Minigene Expression Vectors Containing AE5' And AE3'

Protein C is a factor which plays a critical role in the anti-blood coagulation mechanism. Unlike factor IX, whose level in the circulation substantially increases with advancing age, protein C levels in the circulation do not increase with advancing age, but rather show a slight decrease over time. This decrease in circulating protein C levels is believed by the inventors to be the result of regulation at the gene transcription level. For this reason, the protein C gene provides an interesting exemplary gene for demonstrating the universality of the AE5' and AE3' function in gene expression both *in vitro* and *in vivo* as follows. In this Example, bases are numbered relative to the major transcription start site (+1) as previously described [Miao et al. (1996) J. Biol. Chem. 16:9587-9594].

### A. Construction of Human Protein C Minigene Expression Vectors

The human protein C genomic sequence has been previously reported (GenBank accession number M11228; Figure 12B]. Using this sequence, three protein C minigene expression vectors were prepared. The first human protein C minigene vector (-1426PCm1)

contained the human protein C promoter region of the protein C gene (GenBank accession number M11228; Figure 12 B) ligated to the human protein C cDNA (GenBank accession number X02750; Figure 12 A) which contains the first entire intron and poly-A sequence. The second human protein C minigene (AE5'/-1426PCm1) was the same as the first vector except that it additionally contained the nucleotide sequence AE5' at the 5' end of the human protein C cDNA. The third human protein C minigene (AE5'/-1426PCm1/AE3') was the same as the first vector vector except that it additionally contained the nucleotide sequences AE5' and AE3' at the 5' and 3' ends, respectively, of the human protein C cDNA.

10

15

20

25

30

5

### B. Transient Expression Of human Protein C in vitro In HepG2 Cells

Each of the protein C minigene expression vectors was transfected into HepG2 cells using the FuGene 6 (Boehringer Mannheim). Four to five independent assays of human protein C activity were carried out as previously described [Turkey et al (1999) Throm. Haemost. 81:727-732]. HepG2 cells transfected with the -1426PCm1 vector showed *in vitro* transient protein C activities which were comperable to the activities shown by HepG2 cells transfected with the AE5'/-1426PCm1 vector (*i.e.*, (60-70 ng/10E6 cells/24 hrs).

# C. Generation Of Transgenic Mice Harboring The protein C Minigene Expression Vectors

In order to determine whether AE5' in combination with AE3' is capable of increasing human protein C expression with advancing age, as observed for factor IX expression (Example 3, *supra*), transgenic mice which harbor the protein C minigene expression vectors are generated according to standard methods [Hogan et al. (1994), *supra*]. Briefly, protein C minigene vector plasmids are injected into fertilized eggs of C57B/6 X SJL mice and implanted into foster mother animals (CD-1). Offspring produced are screened for founder animals with high transgene copy numbers (5 -10 copies per genome) using quantitative multiplex PCR analyses of tail tissue DNA samples using two pairs of primers which are designed to amplify a unique fragment from the protein C transgenes and a 494 bp fragment from the mouse β-globin gene, respectively.

Founders are back-crossed with non-transgenic mice (C57B/6 X SJL) to generate F1 progeny animals. Homozygous F2 animals are generated by crossing among heterozygous

F1 littermates and the following generations are similarly generated. Zygosity status of animals is determined by quantitative multiplex PCR analysis as described above. Founder lines for each minigene construct are subjected to longitudinal analysis for their entire life spans up to two years.

Circulatory human protein C levels are monitored during longitudinal analyses of transgenic mice from the representative founder lines carrying the protein C minigene transgenes. Age-regulation patterns of circulatory human protein C levels are compared among all animals for each specific construct, different founder lines, different initial human protein C level at pre-pubertal age (one month) due to transgene positional effects, generation (founders and F1 or F2 progeny), sex, and zygosity (homozygous / heterozygous) status of the transgenes.

Northern blot analyses of the liver RNA samples from animals is carried out using stringent washing conditions to determine whether any changes in circulatory human protein C levels are correlated with similar changes in the steady-state liver human protein C mRNA, rather than with loss of the human protein C transgene with age or with changes in human protein C turnover time. Observation of transgenic animals which contain the human protein C sequence as well as AE5' and AE3' and which increase stable circulatory human protein C levels with increasing animal age, as compared to the levels in transgenic animals which express the human protein C sequence in the absence of AE5' and AE3', demonstrates that the combination of AE5' and AE3' functions in age-stable expression of the exemplary human protein C gene. This observation will confirm that these results may be achieved for genes other than the exempary hFIX and protein C genes.

#### **EXAMPLE 8**

25

30

5

10

15

20

# In Vivo Expression Of Exemplary Expression Vectors Containing The Cytomegalovirus (CMV) Promoter And The AE5' And AE3'

This Example is carried out to demonstrate the universality of the age-related gene expression regulatory function of AE5' and AE3' with viral promoters. The CMV promoter is currently used in several gene therapy constructs but its activity decreases with time in the Liver. Furthermore, the activity of the CMV promoter in the liver of transgenic mice is known to be lower than the activity in other tissues, such as muscle. Thus, this Example

investigates whether the combination of AE5' and AE3' halts or reverses the decline in the activity of the CMVpromoter in the liver.

5

10

15

20

25

30

The above-described -416FIXm1 expression vector (Example 1, and Figure 1) is used to construct a control vector to determine the effect of AE5' and AE3' on liver and circulatory levels of expression of human factor IX in transgenic animals. The control vector in which expression of the hFIX gene is under the control of the CMV promoter in the absence of both AE5' and AE3' is constructed by replacing the human factor IX promoter sequence with the CMV promoter sequence (National Vector Core for Non-Viral Vectors at the University of Michigan) (the CMV promoter is also located between positions +1 to +596 in vector plasmid pCR3 from Invitrogen). The resultant expression vector in which the human factor IX gene is under the control of the CMV promoter is transfected into HepG2 cells. Transfected cells are expected to show human factor IX activity.

The -802FIXm1/1.4 vector of Figure 1 is used to construct a test vector in which the human factor IX promoter sequence of -802FIXm1/1.4 vector (which contains both AE5' and AE3') is replaced with the CMV promoter sequence.

In order to determine whether the combination of AE5' and AE3' is capable of increasing human factor IX expression with advancing age under the control of the CMV promoter, as observed for factor IX expression under the control of the factor IX promoter (Example 3, supra), transgenic mice which harbor either the control vector or the test vector are generated according to standard methods as described supra (Examples 3 and 7). The mRNA levels of human factor IX in the blood, liver and other tissues are monitored during longitudinal analyses of the transgenic mice. Age-regulation patterns of human factor IX mRNA levels in the different tissues are compared among all animals as described supra for each specific construct, different founder lines, different initial human factor IX levels at prepubertal age due to transgene positional effects, generation, sex, and zygosity status of the transgenes.

The observation of transgenic animals which contain the test vector and which increase stable circulatory human factor IX mRNA levels, as compared to the circulatory mRNA levels in transgenic animals which contain the control vector, demonstrates that the combination of AE5' and AE3' functions in increasing the activity of the exemplary CMV promoter.

#### **EXAMPLE 9**

# Liver-Specific Expression Of The Exmplary Human Factor IX Gene Under The Control Of The CMV Promoter

This Example investigates whether the presence of AE5' imparts liver specific activity to the CMV promoter, which otherwise drives gene expression in several tissues in addition to the liver.

The -802FIXm1 vector which contains AE5' and lacks AE3' (Figure 1) is used to construct a test vector in which the human factor IX promoter sequence of the -802FIXm1 vector is replaced with the CMV promoter sequence. This test vector is used in parallel experiments with the control vector of Example 8 in which expression of the hFIX gene is under the control of the CMV promoter in the absence of both AE5' and AE3'. Northern blot analysis is carried out as described *supra* (Example 3) in transgenic mice carrying the control vector or test vector. Animals expressing hFIX at high level are sacrificed at one month of age and total RNA is extracted from liver, lung, intestine, muscle, kidney, brain and heart and from untransfected HepG2 cells (negative control). The levels of hFIX mRNA in the different tissues are compared. It is expected that transgenic animals harboring the control vector will express hFIX mRNA in liver as well as in at least one other tissue. In contrast, the observation that transgenic animals which harbor the test vector express hFIX mRNA in the liver and not in other tissues indicates that AE5' confers liver-specific activity to the exemplary CMV promoter.

From the above, it is clear that the invention provides methods for age-related and liver-specific gene expression and models for age-related and liver-specific diseases.

25

30

5

10

15

20

### **EXAMPLE 10**

# Construction Of A Series of Exemplary Human Protein C Minigene Expression Vectors

This Example and the following Examples 11-12 were carried out to further demonstrate the universality of the function of AE5' and AE3" sequences with respect to regulating expression of the exemplary human protein C as described in Example 7, *supra*. The human protein C genomic sequence has been previously reported (GenBank accession

number M11228; Figure 12B]. In particular, Figure 14 shows the nucleotide sequence (SEQ ID NO:85) of the 5'-end of the human protein C gene [Miao et al. (1996) J. Biol. Chem. 16:9587-9594]. In Figure 14, bases are numbered relative to the major transcription start site (+1) as previously described [Miao et al. (1996) J. Biol. Chem. 16:9587-9594]. Two minor start sites are marked with double asterisks. Exons are underlined. The translation start codon (ATG) is shown in boldface.

5

10

15

20

25

30

Using this sequence, eight protein C minigene expression vectors were prepared as shown in Figure 15. Fat lines at 5' and 3' ends represent flanking sequences. Introns are shown by blank rectangles. Exons are shown by shaded rectangles (corresponding to the 3' UTR part of the last exon). Arrows indicate transcriptional start sites. The vectors contained the human FIX AE5' sequence (SEQ ID NO:1; nt -802 through -771 of Figure 8; 32 bp in size) linked to Sph I linker sequences and/or the human FIX AE3'' sequence (SEQ ID NO:93; nt 32,110 through nt 32,263 of Figure 8, 154-bp long) linked to Sse8387 I linker sequences. All the amplified sequences of the human protein C minigene vectors were verified by dideoxy sequencing.

The first human protein C (hPC) minigene vector (-1462hPCm1) was composed of the 5' flanking sequence of human protein C up to nt -1462, exon I sequence, the complete first intron at the natural site (1431 bp in length) and the contiguous following sequence containing exons 2-9 (derived from the hPC cDNA) and the 3' immediate flanking genomic sequence through nt 11,108 (325 bp in length). As shown in Figure 15, nucleotide 67 was the first base of intron I, nucleotide 1,497 was the first base of exon 2 (the first base of the Met codon is nucleotide 1,514), and nucleotide 10,488 was the last base of the stop codon.

To construct -1462hPCm1, a region spanning nt -1462 through 1,560 (3,022 bp in length) of the hPC gene was amplified by PCR (Expand High Fidelity PCR System, Boehringer Mannheim) using 5' and 3' primers containing Sph I linker and the unique internal Msc I site in the exon 2, respectively, and human genomic DNA as the template. The generated fragment, containing the 5' flanking sequence, exon I, intron I and a short 5' portion of exon 2 to the internal Msc I site, was then inserted into a hPC cDNA plasmid, PUC119-hPC, in between Sph I and Msc I sites by replacing its 5' portion of the hPC cDNA sequence. The 3' end of the resulting minigene, -1462hPCm', contained the entire 3' UTR, but only up to poly(A) attachment site (nt 10,783). Its 3' end region (the 3' sequence beyond the internal Sse8387 I site in the 3' UTR) was then freed by Sse8387 I/EcoR I

double digestion, and replaced with a Sse8387 I/EcoR I fragment (612 bp in the length, spanning nt 10,497 through 11,108 of hPC gene), which was generated by PCR. The hPC minigene, thus constructed was named -1462hPCm1, (approximately 4981 bp in length) and served as a parent construct for generating other hPC minigene constructs for generating other hPC minigene constructs as described below. This hPC minigene vector was used to construct the first transgenic mouse colony, which was used as a positive control for hPC expression.

5

10

15

20

25

30

The second human protein C minigene vector (-82hPCm1) was another control human hPC minigene. It was the same as -1462hPCm1, except that its promoter region was extended only up to nt -82 instead of to nt -1462, and thus contained the hPC upstream sequence (SEQ ID NO:86) from nt -82 to nt +1. Minigene -1462hPCm1 was subjected to Sph I digestion, followed by partial MscI digestion, releasing the 5' half region spanning nt -1462 in the 5' upstream through nt 1,547 at the internal Msc I site in exon 2. Due to another Msc I site in the first intron, partial digestion was needed to get the needed 5' end half fragment. This region was then replaced by smaller Sph I/Msc I fragments generated by PCR, spanning nt -82 in the 5' upstream through the internal Msc I site, thus generating -82hPCm1.

The third human protein C minigene vector (AE5'/-1462hPCm1) was the same as -1462hPCm1, except that it had an AE5' sequence (32 bp in length) inserted at the 5' end at an Sph I site.

The fourth human protein C minigene vector (-1462hPCm1/AE3'') was the same as -1462hPCm1, except it had an AE3'' sequence inserted at the Sse8387 I site within the 3' untranslated region (UTR) of the hPC minigene.

The fifth human protein C minigene vector (AE5'/-1462hPCm1/AE3'') was the same as -1462hPCm1, except it had both AE5' and AE3''. The AE5' fragment (32 bp in length, spanning nt -802 through -771 of hFIX gene) was amplified by PCR (Expand High Fidelity PCR System, Boehringer Mannheim) with human genomic DNA as template and PCR primers containing Sph I linker sequences. The AE3'' fragment spanning nt 32,110 through nt 32,263, and containing the potential stem-loop structure forming region (nt 32,142 through 32,243) was produced by PCR using primers containing the internal Sse8387 I site sequences. AE5' and AE3'' fragments with Sph I or Sse8387 I sticky ends, respectively, were then inserted into -1462hPCm1 at Sph I site at the 5' end and Sse8387 I site in the 3'

UTR, respectively, thus generating minigene AE5'/-1462hPCm1/AE3''. Therefore, AE5'/-1462hPCm1/AE3'' is the same as -1462hPCm1 except it has both AE5' and AE3' sequences of the hFIX gene at the 5' end Sph I site and Sse8387 I site in the 3' UTR, respectively.

The sixth and seventh human protein C minigene vectors (-849hPCm1 and -802hPCm1) were the same as -1462hPCm1, except that their 5' end sequences extended to nt -849 and -802, respectively, instead of to -1462. Construction of these minigenes was essentially the same as construction of the second human protein C minigene (-82hPCm1); minigene -1462hPCm1 was subjected to Sph I digestion, followed by partial MscI digestion, releasing the 5' half region spanning nt -1462 in the 5' upstream region through nt 1,547 at the internal Msc I site in exon 2. Due to another Msc I site in the first intron, partial digestion was needed to get the needed 5' end half fragment. This region was then replaced by smaller Sph I/Msc I fragments generated by PCR, spanning nt -849, or -802 in the 5' upstream through the internal Msc I site, thus generating -849hPCm1 and -802hPCm1, respectively.

The eighth human protein C minigene vector (AE5'/-82hPCm1) was generated by inserting the AE5' sequence with Sph I sticky ends into the second vector, -82hPCm1, at the 5' end Sph I site.

20 EXAMPLE 11

5

10

15

25

30

# Transient Expression of Exemplary Human Protein C Minigene Expression Vectors In Vitro In Human Hepatoma HepG2 Cell Line

Transient expression activities of the hPC minigenes of Example 10 were tested with HepG2 cells as described in Example 2, *supra*. Figure 16 shows the relative *in vitro* transient expression activities and standard deviations of the human protein C minigene expression constructs relative to the activity (approximately 70 ng/10<sup>6</sup> cells/48 hr, defined as 100% activity) of the control -1462hPCm1 construct.

The relative transient expression activities in Figure 16 show that the results obtained with AE3" when in tandem with the hPC gene are fully consistent with what the inventors previously observed with AE3" when in tandem with the hFIX gene (Figure 1); the presence of AE3" showed approximately a 30% suppression in transient expression in comparison to the minigenes which lacked AE3" (i.e., -1462hPCm1/AE3" compared with

-1462hPCm1, and AE5'/-1462hPCm1/AE3'' compared with AE5'/-1462hPCm1). The control constructs -1462hPCm1 and -82hPCm1 showed similar transient expression activities to each other.

As discussed above, these results were surprising because they were contrary to those previously reported by Miao et al. (1996), *supra*, when using a heterologous reporter gene, chloramphenical acetyltransferase (CAT), under the transcriptional control of varying lengths of the protein C 5'-end sequences.

5

10

15

20

25

30

#### **EXAMPLE 12**

### Generation And Analysis Of Transgenic Mice Harboring Human Protein C Minigene Expression Vectors

Transgenic animals were constructed using each of the eight expression plasmids described above in Example 10 according to standard methods [Hogan et al. (1994) in "Manipulating the Mouse Embryo, a Laboratory Manual" (Cold Spring Harbor Press, New York, 2nd Edition) as described in Example 3, *supra*. Fertilized eggs of mice were microinjected with the minigene transgene DNA and implanted into foster mother animals.

Circulatory hPC levels were monitored during longitudinal analyses of transgenic mice harboring the -1462hPCm1, -82hPCm1, and AE5'/-1462hPCm1/AE3'' minigene transgenes. At various ages, starting at one month of age, transgenic mice were individually subjected to blood sample collection via tail-tip snipping, and the obtained serum was routinely used to quantify hPC levels in the circulation using ELISA for each age point. The ELISA assay employed a mouse monoclonal anti-hPC antibody (Celsus Laboratories) as a first antibody and a rabbit polyclonal anti-hPC antibody (Celsus Laboratories) as a second antibody. Pooled human plasma (George King Bio-Medical) was used to prepare a hPC standard curve for each assay. The results are shown in Figure 17. The labeling in Figure 17 reflects the tag numbers of animals containing each minigene construct. Figure 17 shows representative animals with -1462hPCm1 (A), -82hPCm1 (B), and AE5'/-1462hPCm1/AE3'' (C) expression vectors.

Importantly, Figure 17 shows that age-regulation patterns were remarkably similar among all animals for each specific construct. In particular, the results show that transgenic animals containing the -1462hPC m1 construct contained age-stable levels of human protein C, *i.e.*, the animals expressed relatively constant levels of human protein C at different time

points during the life span of the transgenic animals (Figure 17A). In direct contrast, the presence of AE5' and AE3' sequences resulted in increased expression levels of human protein C over time (Figure 17C). These results confirm the universality of the function of AE5' and AE3'' sequences in regulating expression of operably linked genes in an age-related manner.

5

10

15

The data also shows that, whereas transgenic animals containing the -1462hPC m1 construct exhibited relatively constant and relatively high levels (from about 100 to about 3000 ng/ml) of human protein C over time (Figure 17A), in dramatic contrast, transgenic animals containing the -82hPCm1 construct exhibited relatively low levels (from about 5 to about 40 ng/ml at 1 month of age) of human protein C which declined at a precipitous rate over time. Indeed, by the age of 5 months, human protein C levels were undetectable in all transgenic animals harboring the -82hPCm1 construct. These results demonstrate that the nucleotide sequence from nt -1462 to nt -83 of the human protein C gene directs age-stable expression as well as relatively higher levels of expression (as compared to the levels in the absence of the nucleotide sequence from nt -1462 to nt -83) of operably linked sequences of interest.

#### **CLAIMS**

1. A recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3.

### 2. A method, comprising:

5

10

15

- a) providing: i) a cell, ii) a nucleic acid sequence of interest, iii) a promoter sequence, and iv) one or more age regulatory sequences selected from SEQ ID NO:1, SEQ ID NO:3, a portion of SEQ ID NO:1, and a portion of SEQ ID NO:3;
- b) operably linking said nucleic acid sequence of interest, said promoter sequence, and said one or more age regulatory sequences to produce a transgene; and
- c) introducing said transgene into said cell to create a treated cell under conditions such that said nucleic acid sequence of interest is expressed in said treated cell.
- 3. A substantially purified nucleic acid sequence comprising at least a portion of SEQ ID NO:93.
- 20 4. The nucleic acid sequence of Claim 1, wherein said portion has age-related regulatory activity.
- 5. The nucleic acid sequence of Claim 1, wherein said portion is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:137, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO

WO 00/75279



NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144.

6. The nucleic acid sequence of Claim 3, wherein said portion is SEQ ID NO:91.

5

7. A recombinant expression vector comprising in operable combination i) a nucleic acid sequence of interest, ii) a promoter sequence, and iii) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof.

10

8. The expression vector of Claim 7, , wherein said nucleic acid sequence of interest encodes a protein selected from factor VIII, factor VII, factor IX, factor X, prothrombin, protein C, antithrombin III, tissue factor pathway inhibitor, LDL-receptor, human α1-antitrypsin, antithrombin III, PEA-3 protein, β-galactosidase, and luciferase.

15

9. The expression vector of Claim 7, wherein said promoter sequence is selected from human factor IX promoter, cytomegalovirus promoter, tRNA promoter, 5S rRNA promoters, histone gene promoters, RSV promoter, retrovirus LTR promoter, SV40 promoter, PEPCK promoter, MT promoter, SRα promoter, P450 family promoters, GAL7 promoter, T<sub>3</sub> promoter, SP6 promoter, K11 promoter, and HIV promoter.

20

25

30

10. The expression vector of Claim 7, , wherein said portion of SEQ ID NO:93 is selected from SEQ ID NO:91, SEQ ID NO:94, SEQ ID NO:95; SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134, SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, and SEQ ID NO:144.



- 11. The expression vector of Claim 10, , wherein said portion is SEQ ID NO:91.
- 12. The expression vector of Claim 7, further comprising in operable combination an age-related regulatory sequence selected from SEQ ID NO:1 and portions thereof.
  - 13. A host cell containing the recombinant expression vector of Claim 7.
- 14. The host cell of Claim 13, wherein said host cell is comprised in a tissue or organ in a living animal.
  - 15. The host cell of Claim 13, wherein said host cell is a gamete.
- 16. The host cell of Claim 13, wherein said host cell is selected from bacterial cell, yeast cell, plant cell, insect cell, and mammalian cell.
  - 17. A method for expressing a nucleic acid sequence of interest, comprising:
  - a) providing:
    - i) a cell;
    - ii) a nucleic acid sequence of interest;
    - iii) a promoter sequence; and
- iv) an age-related regulatory sequence selected from SEQ ID NO:93 and portions thereof;
- b) operably linking said nucleic acid sequence of interest, said promoter sequence, and said age-related regulatory sequence to produce a transgene; and
- c) introducing said transgene into said cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in said treated cell.
- 18. The method of Claim 17, wherein said treated cell is comprised in a tissue or organ in a living animal.

5

10

15

20

25

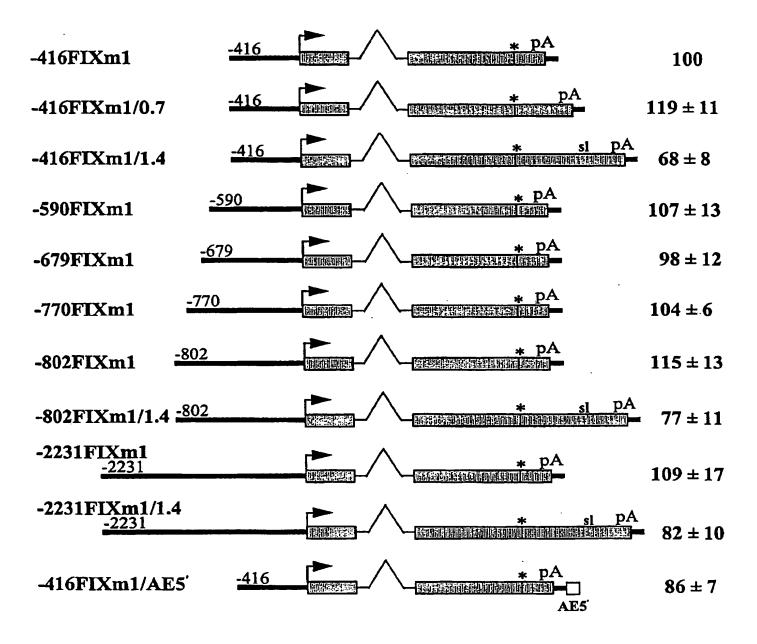
5

10

15

- 19. A substantially purified nucleic acid sequence comprising a nucleotide sequence selected from at least a portion of SEQ ID NO:85, and at least a portion of SEQ ID NO:92.
  - 20. A method for expressing a nucleic acid sequence of interest, comprising:
  - a) providing:
    - i) a cell;
    - ii) a nucleic acid sequence of interest;
    - iii) a promoter sequence; and
- iv) a nucleotide sequence having activity selected from age-related regulatory activity and regulatory activity, said nucleotide sequence selected from SEQ ID NO:92, a portion of SEQ ID NO:92, SEQ ID NO:85, a portion of SEQ ID NO:85, SEQ ID NO:89, and SEQ ID NO:90;
- b) operably linking said nucleic acid sequence of interest, said promoter sequence, and said nucleotide sequence to produce a transgene; and
- c) introducing said transgene into said cell to create a treated cell under conditions such that the nucleic acid sequence of interest is expressed in said treated cell.

Figure 1



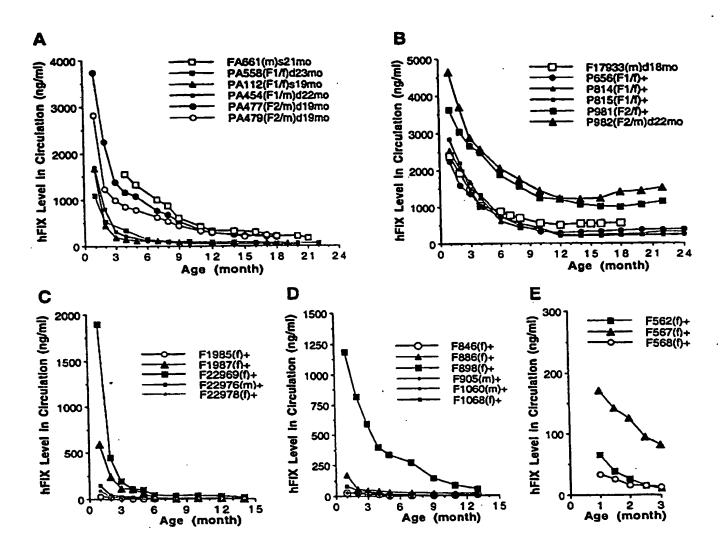
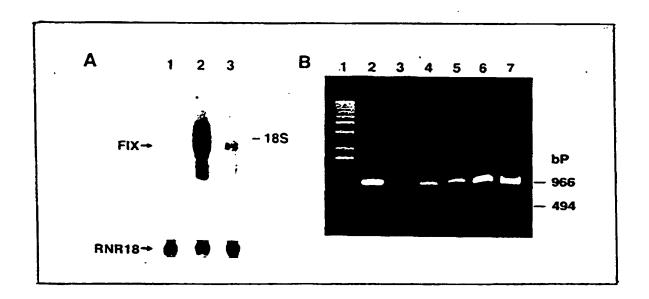


Figure 2

Figure 3



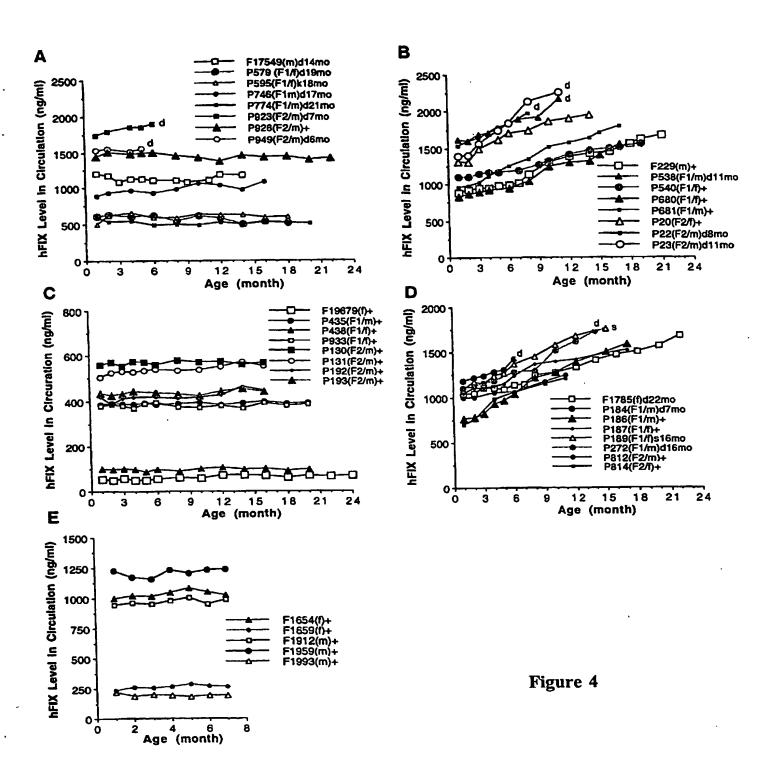


Figure 5

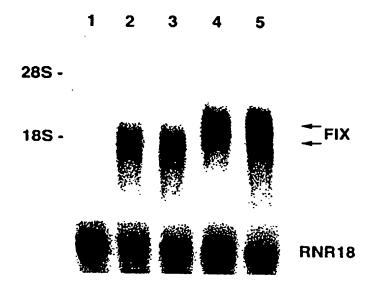


Figure 6

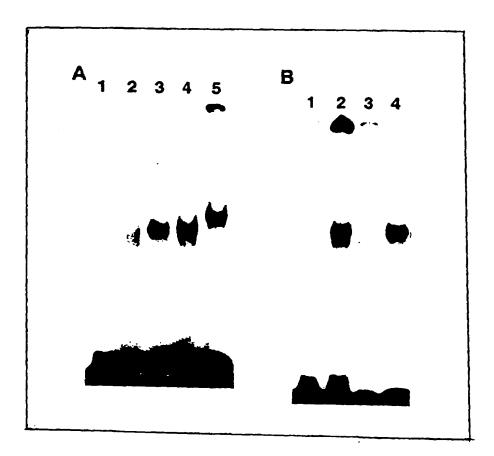
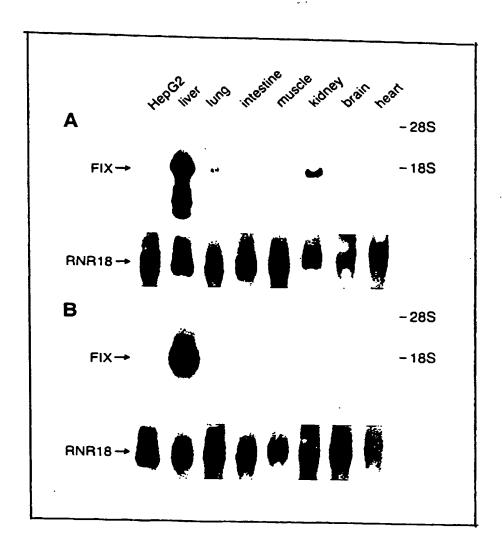


Figure 7



### Figure 8A

	WWCCCCAT	TGTCTCATTC	CAAAATCACC	TTANGATGGA	TAGGCAACTT	CAGCANAGEC	TCAGGATAAC	AAAATCAATG	TGC23433TC	-2866
ACAGGGATTC T	TATALACCA	ATACCAGACA	AACAGAGAGG				COMMONTAL	~ · · · · · · · · · · · · · · · · · · ·		-2766
CIUCIIVON O	MANUTURE.	GRACUTURE	AACCACALCT			141116166	ATICABACA:	100001011		-2666
										-2566
AGAATTOGAA A	MAYETYCE!	TAAAGTTCAT	ATGGAACCAA	AAAAGAGCCC	GCATCGCCAA	GTCAATCCTA	AGCCAAAAGA	ACAAAGCTGG	AGGCATCATG	-2466
										-2366
CAGAAATAAT G	CCACATATC	TACAACTATC	TGATCTTTGA	AAAACCTGAC	<b>AAAAACAAGA</b>	aatggggaaa	GGAATCCCTA	TAATAATA	GGTGCTGGGA	-2266
										-2166
										-2066
										-1966
AATTTTTGCA A GGGTGAAGGA T										-1866
										-1766
										-1666
ATGACTTGGA A	ACCANCCCAA	ATGTCCAACA	ATGATAGACT	GGATTAAGAA	AATGTGGCAC	ATATACACCT	AGGRATACTA	GCCACCCATA	CACAATAGCA	-1466
										-1366
AGGTGGGAAC T	<b>rgaacaatga</b>	GAACACTTGG	ACACAGGAAG	GGGAAÇATCA	CACACCGGGG	CCTGTTGTGG	GGTGGGGGGC	GAGGGGAGGG	ATAGCATTAG	-1266
										-1166 -1066
										-966
										-866
										-766
										-666
AAATGGAGCA T AATGTGATCC A	TAAAACAAA	AACAAATAAA	CTCCACTGTAX	AACATTTAGA	AAAAAATAGG	AAAAAAACTA	TCAGGATCTA	GTGTTAGTGG	AAGAGTTCTA	-566
AGGCTGGAGA C		AATCCACCTA	CIGGACTACA	CLERCHAN	AAAATTCTAC	TCTGTGAAAG	ACCTAATTAA	GAGGACAAAA	GACAAGCTAC	-466
										-366
GATGAACTGT G	CTGCCACAG	TARATGTAGC	CACTATGCCT	ATCTCATAC	TGALCATOR	TCACCTCCTC	CTCTCTGACA	AAGATACGGT	GGGTCCCACT	-266
AGCCCACGAA A	ATCAGAGGTG	AAATTTAATA	ATGACCACTG	CCCATTCTCT	ICACATICICO	CAACAGGCCA	TTCCLARTIC	ACCANACIOCA .	CAGTGGCAGA	-166
					1040100	CHICKOCCA	***********	reconstance		-66
						_		•	-45	
GATGGACATT A	ATTTCCCAGA	AGTAAATACA	<b>GCTCAGCTTG</b>	TACTTTGGTA	CAACTAATCG	ACCTTACCAC	TTTCACAATC	TECTACCAAA	Net	32
				•						32
Gln Arg Val	Asn Net II	e Met Ala (	Slu Ser Pro	Gly Leu Ile	The Ile C	ys Leu Leu (	ly Tyr Leu	Leu Ser Ala	Glu Cvs	
cua cae ara	AAC ATG AT	nc atg gca (	SAA TCA CCA	GGC CTC ATY	C ACC ATC T	GC CTT TTA	GÁ TẤT CTA	CTC AGT GCT	GAA TGT	113
-10		•	•							
Thr										
ACA G GTTT G	STITECTATA	TTAAAATACA	TTGAGTATGC	TTGCCTTTTA	GATATAGAAA	TATCTGATGC	101CIICIIC	ACTAAATTT	GATTACATGA	211
PERCACAGOA A	********	CECERACACO	C10010010							
TTTGACAGCA A	ATGCTTATGA	TOCATTATO	CAUCACGCAG	CTTCCTAACT	ACTGGTTCTT	TGTTAGCTAG	GITTICITCI	TCTTCATTTT	TAAAACTAAA	
TAXAAAATT A	AAAAGTGGGA	ABACABAGAA	ATACCACALE	ACTOTICAGE.	TEATGATTTG	GICATOTAAT	TCCTGTTAGA	AAACATICAT	CTCCTTGGTT	411
										\$11
										611 721
	VVV1/V1/1/1/	GILILIATET	CARACATOTT	CCACTTCAT	TATTTCCCCA	******	ALT CAME FOR	CCCC LCCCC		811
										911
CATACAGAM W		CCATTTTGGAC	AAACAGCATG	TTCTCACAGG	BACCATTAN					
TGACAGTACC A					<b>WUACULITYI</b>	CACACTTACT		CTAGAATCAA	ATCTACTACC	1011
		CTGCCAACCC	TAAGCACCCC	CAGAAAGCTG	ACTRICCCT	TOTALCOLAC	TOTCAACTIT	ATCTCACCOC		1011 1111
CICAAAIGCI G	GAAATAACGA	TARRARA	TARGUACCCC	CAGAAAGCTG	ACTGGCCCTG	TGGTTCCCAC	TCTCAACTTT	ATGTCAGCTG	TGAAATCAGA	1111
AGGACATAAA G	GAAATAACGA GCAAGGCCAT	TARARARAT	TARGEACCCC	AAACTAGCAA GACATCTGGG	ACTGGCCCTG AGTGAGTAAA AGGACAAAGC	GTCAAGGGAT	TCCAGACATG AAAGAAAATT	ATGTCAGCTG	TGARATCAGA	1111 1211
AGGACATAAA G	GAAATAACGA GCAAGGCCAT GCCTTCTGCC	TAMAMAMA TAGATATATC CCCTTGAAGA	TAGEACCCC TACAGAGGTT TCATTAGTGT CTTCAGATGC	AAACTAGCAA GACATCTGGG TGGGGAAAGG	ACTGGCCCTG AGTGAGTAAA AGGACAAAGC ATAGATAAGA	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC	TOTCACTIT TCCAGACATG AAAGAAAATT TTTCTTCTAT	ATGTCAGCTG TGTTGGAAAA ATAAGTGGTG	TGANATCAGA CTCACAAAGC AGATGATGAA	1111 1211
AGGACATAAA G GGTTGTAAGA G CTAGCGATAA A	EAAATAACGA ECAAGGCCAT EGCTTCTGCC ACCTGAAGGG	TANAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA	TANGEACECE TACAGAGGTT TCATTAGTGT CTTCAGATGC TACGATCCC	AAACTAGCAA GACATCTGGG TGGGGAAAGG	ACTGGCCCTG AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG	GTCAAGGGAT ATCCAAACCC ATAAGGATGA	TGTCAACTIT TCCAGACATG AAAGAAAATT TTTCTTCTAT ACCTGGCTTT TCTCCAGCCA	ATGTCAGCTG TGTTGGAAAA ATAAGTGGTG TGGAGCCTGG	TGAAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA	1111 1211
AGGACATAAA G GGTTGTAAGA G CTAGCGATAA A CCCTTGAGGA A	GAATAACGA GCAAGGCCAT GCCTTCTGCC ACCTGAAGGG AGGGGCCAGG	TANAMANA TAGATATATC CCCTTGAAGA AAGTTAAGTA	TANGEACECE TACAGAGGTT TCATTAGTGT CTTCAGATGC TACGATCCCC TACGATCCCC	CAGANAGCTG ANACTAGCAA GACATCTGGG TGGGGAAAGG AGATAATACT	ACTGGCCCTG AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG	GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT	TGTCAACTIT TCCAGACATG AAAGAAAATT TTTCTTCTAT ACCTGGCTTT TCTGCAGCCA	ATGTCAGCTG TGTTGGAAAA ATAAGTGGTG TGGAGCCTGG TTGTAGCCAG	TGAAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AGATAATAAG	2111 1211 2311 1411 1511 1611
AGGACATAAA G GGTTGTAAGA G CTAGCGATAA A CCCTTGAGGA A CGGTGAGTGA T	GAAATAACGA GAAAGACCAT GCAAGGCCAT GCCTTCTGCC ACCTGAAGGG AGGGGCCAGG ETTGCTGAGA	TAAAAAAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATT	TAAGLACCCC TAAGAGGTT TCATTAGGGTC CTTCAGATCCC TAAGGATAGA TTCATGCTCC	CAGAMAGCTG AMACTAGCAA GACATCTGGG TGGGGAAAGG AGATAATACT CAGTATAAAT TGCCTTTAGG	ACTGCCCTG AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT	TGTCAACTIT TCCAGACATG AAAGAAAATT TTTCTTCTAT ACCTGGCTTT TCTGCAGCCA TAAACTCTCA	ATGTCAGCTG TGTTGGAAAA ATAAGTGGTG TGGAGCCTGG TTGTAGCCAG TTGGCTTCTA	TGAAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AGATAATAAG AAAGGAGTTT	2111 1211 2311 1411 1511 1611 1711
AGGACATAAA G GETTGTAAGA G CTAGCGATAA A CCCTTGAGGA A CGGTGAGTGA I CAGTTAGTT I	GAAATAACGA GAAAGGCCAT GCAAGGCCAT GCCTGAAGGG ACCTGAAGGG AGGGGCCAGG ITTGCTGAGA ITGTAAAGTG	TAAAAAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATT TATGCATCAA	TANGEACECC TACAGAGGTT TCATTAGTGT CTTCAGATGC TACGATCCCC TAAGGATAGA TTCATGCTGC AGATGCCTT	CAGAMAGETG AMACTAGEAA GACATCTGGG TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA	ACTGGCCCTG AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGCT	TOGTTCCCAC GTCAAGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTCG	TGTCAACTTT TCCAGACATG AAAGAAAATT TTTCTTCTAT ACCTGGCTTT TCTGCAGCCA TAAACTCTCA AATTTTGAAA TGCCAGCTAAA	ATGTCAGCTG TGTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGTAGCCAG TTGGGTTCTA TTAAAACAGT	TGARATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AGATAATAAG AAAGGAGTTT TCTGTAAAAC	2111 1211 2311 1411 1511 1611 1711 1811
AGRACATARA G GGTTGTRAGA G CTAGCGATAR A CCCTTGAGGA A CGGTGAGTGA T CAGTTTAGTT T ATGGGGATAR A	SAAATAACGA SCAAGGCCAT SGCTTCTGCC ACCTGAAGGG AGGGGCCAGG FTTGCTGAAGTG TTGTAAAGTG ACCAGACTCC	TANANAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATT TATGCATCAA CTCTTTGATC	TANGENCECC TANGENGETT TCATTAGTGT CTTCAGATGC TANGGATAGA TANGGATAGA TTCATGCTGC AGATGCCCT TANGGATAGA TANAGCAGCCT TANGGATAGA	CAGAAAGCTG AAACTAGCAA GACATCTGGG TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGGAGG	ACTGGCCCTG AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGGTT TGAGGAGGTT	TGGTTCCCAC GTCAAGGGAT ATCCAAACCE ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTCG ACAACTACGG	TGTCANCTTT TCCAGACATG AMAGMAATT TTTCTTCTAT ACCTGGGTTT TCTGCAGCCA TAMACTCTCA AATTITGAAA TGGCAGGTAC	ATCTCACCTG TCTTGGAAAA ATAACTGCTG TGGAGCCTGG TTGTAGCCAG TTGGACCTTCTA TTAAAACAGT TCTGTCAGGG	TGARATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AGATAATAAG AAAGGAGTTT TCTGTAAAAC TACTAGGGGT	2111 1211 2311 1411 1511 1611 1711 1811 1911
AGGACATANA G GGTTGTANGA G CTAGCGATAN A CCCTTGAGGA A CCGTGAGTGA T CAGTTTAGTT T ATGGGGATAN A ATAGAGANAG G CTGATATANAG G	SAATAACGA SCAAGGCLAT SCETTCTGCC ACCTGAAGGG AGGGGCEAGG FTTGCTGAGA FTGTAAAGTG ACCAGACTCC GAACACATTA GGCATTTTAT	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATT TATGCATCAA CTCTTTGATC AATGGGGAAA GCAAAGAAGA	TANGUACCCC TACAGAGGTT TCATTAGTGT CTTCAGATGC TACGATCCCC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAAGCAGCA CAATTGATAG ATCATCCTG	CAGAAAGCTG AAACTAAGCAA GACATCTGGG TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAGAATA	ACTGGCCCTG AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TGAGAGGTTT TTACTGCAGG	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTCG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTCGCAC	TOTCAGENTE TCCAGACATE ANAGMANTI TITCTTCTAT ACCTGGCTIT TCTGCAGCCA TANACTCTCA ANTITTGANA TGGCAGGIAC GATAMATGT GATAGAGTAGG	ATGTCAGCTG TGTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGTAGCCAG TTGAAACAGT TTAAAACAGT TGTGTCAGGG GCACTAGGTA CGTTCTCT	TGANATCAGA CTCACAAAGC AGATGATAATGA GAMATAATGA AGATAATAAG ANAGGAGTTT TCTGANAAC TACTAGGGGT CTANAGGATC CTTANATGT	1111 1211 1211 1411 1511 1611 1711 1811 1911
AGGACATAAA G GGTTGTAAGA G GCTAGCGATAA A CCCTTGAGGA T CGGTGAGTGA T CAGTTTAGTT T ATAGAGATAA A ATAGAGAAAC C CTGATATAAG G ACTATGACAA	SAMATAACGA SCAAGGCCAT SCCTTATGCC ACCTGAAGGG AGGGGCCAGG FITGCTGAGA FITGTAAAGTG ACCAGACTCC SAACACATTTA GGGATTTTA GTGAGACAGG	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATT TATGCATCAA CCCATTGATC AATGGGGAAA GCAAAGAAGG	TAGAGGTT TACAGAGGTT TCATTAGTGT CTTCAGATGC TACAGATCCCC TAAGGATAGATGCTCT TAAAGCAGCA AGATGTCCTT TAAAGCAGCA CAATTGATAG ATCACTCGTG AAGATGATAG AAGAGCAGCA AAGAGCAGCA AAGAGCAGCA AAGAGCAGGTG	CAGAAAGCTG AAACTAGCAA GACATCTGGG TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGC CATTCAGCA TGAGGCCAGG AGAGAATA AAAGACTCA ATCAGATAAT	ACTGECCTTE AGTGAGTANA AGGACANGC ATAGATANGA AAGGAGANAG GCAGCACTCT TTATTATTGC TTACTGAGTTT TTGAGAGGTTT TTTCATCTG GCATTTCANG GCAGTTATTA	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT ACAGTTTGG ACAACTACGG CAATATAGG GGTCTTAAAA AACGTGAGGT	TGTCAACTIT TCCAGACATC AAAGAAAATT TTTCTTCTAT ACTGGCTT TCTGCAGCCA AAAATTTTGAAA TGGCAGGTAC GATAAAATGT GATAAAATGT GATAAAATGT GATAAAATGT ATTCAGGAG ATTTCAGGAG	ATCTCAGCTG TOTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGTAGCCAG TTGTAGCCAG TTGAAAACAGT TGTGTCAGGG CCACTAGGTA CGTTCTCTCT TTTGTTAGCTAG TAAAACAGT	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTTT TCTGTAAAAC TACIAGGGGT CTTAAGGGATC CTTAAGGGATC CTTAAATGT TCCATATGG	1111 1211 1311 1411 1511 1611 1711 1811 1911 2011
AGDACATAAA G GGTTGTAAGA G GCTTGTAAGA G CCATGAGGA A CCGTGAGTGA T CAGTTTACTT I ANGGGATAA A ATAGAGAAAA G CTGATATAAG G ACTATGACAA G ACTATGACAA G CGAGGTAAA A	GAAATAACGA GCAAGGCCAT GCCTTCTGCCGA ACCTGAAGGG AGGGGCCAGG FTTGCTGAGA FTTGTAAAGTG ACCAGACTCC GAACACATTA GGCATTTTAT GTGAGACAGG BTGAGACAGG	TAMAMAAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TGTTTGCATT TATGCATC TATGCATC AATGGGGAAA GCAAAGAAGG TAAACTAGG	TAGAGGTT TACAGAGGTT TCATAGGTT TAGAGTCCC TAGGATCCC TAAGGATAGA TTCATGCTGC TAAAGCAGCT TAAAGCAGCA TACACTCGTG TAAAGCAGCA TACACTCGTG AGAGGTTGTC	CAGAAAGCTG AAACTAGCAA GACATCTGGG TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAAGACATA ACAAGACTAC ATCAGATAAT ACAAGACTAC ATCAGATAAT ACAAGACTAC ACAAGACTAC ACAAGACTAC ACAAGACTAC ACCACCCCAC ACCACCCCAC ACCACCCCCAC ACCACC	ACTGECCTTC AGTGGTAAA AGGACAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TTTCATCTG GCTTTGCAAG GAAGTCATAA	TGGTTCCCAC TGCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT ACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTAAGGAG CCCTAAGGAG	TOTCANCTITI TOCAGACATE ANAGAMANT TITCTTCTAT ACCTGGCTT TCTGCAGCCA TANACTCTCA ATTTTGAA TGCCAGGTAC GATAMANTGT GATAMANTGT GATAGAGTAGG ATTTCAGGAG ATTGAGCAAT TGCAGCACAT TGCACACAT TGCACACA	ATCTCAGCTG TOTTGGAMA ATAMCTCGTG TGGAGCCTGG TTGGGCTTCTA TTAMACAGT TTGTGTCAGG GCACTAGGTA CGTTCTCTC TTTTGTATGG AMATGCAAT	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA AAAGGAGTTT TCTGTAAAAG TACTAGGGGT CTTAAAGGATC TTCCATATGG ATGGAGGTAT	1111 1211 1311 1411 1511 1611 1711 1811 1911 2011 2211
CTAMATICAL AGGACATANA GOTTOTNAGA GOTTOTNAGA CCOTTGAGGA CCOTTGAGGA CGOTGACTGA ATAGGGATAN ATAGGGATAN ATAGAGANA CTATATAAG ACTATGACAN CGAMOTATAN CTAGAGAGAA CCTAGGGAGAA ACTATGACAN CCTAGGAGAAA CCTAGGAGGAA ACTATGACAN CCTAGGAGGAA	GAMATAACGA GCAAGGCCAT GCAAGGCCAT ACCTGAAGGG ACGGGCCAGG ITTGCTGAAGT ITGTAAAGTG ACCAGACTCC GAACACATTA GGGCATTTTAT GTGAGACACACACACACACACACACACACACACACACACA	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATTT TATGCATTA CTCTTTGATC AATGGGAAA CTCATTGATC TACCACTGAT TACCACTGAT TACCACTGAT GACTATTGCA	TAGAGGTT TACTAGAGGT TCATTAGTGT CTTCAGATGC TACGATCCCC TAAGGATAGA TTCATGCTGC TAAAGCAGCA AGATGCATC TAAAGCAGCA AATTGATAG AGGCTGGTC GGCTGATTTAA	CAGAAAGCTG AAACTAGCAA GACATCTGGG TGGGGAAAGG TGAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAGAATA ATCAGATAAT ATCAGATAAT GGATGCCAG GGAGAGATAA	ACTGECCTTE AGGACALAGE AGGACALAGE ATAGATAAGA AAGGAGALAG GCAGCACTCT TTATTATTAGC TTACTGCAGT GCTTTCATCTG GCTTTCATCAGT TCTGGCAACA AGGCTACAACA AGGCTACACA	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTCAGGT ACCTAAGGAG CCGTAAGGAG	TGTCAACTIT TCCAGGAAAATT TTTCTTCTAT ACCTGGCTT TCTGCAGCCA AAATTTCAAA TGCCAGGTAC GATAAAATGT GATAGAGTAG ATTTCAGGAA ATTTCAGGAA ATTTCAGGAAA ATTAGACAAT AGGATACTCA	ATGTEAGGTG TGTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGTAGCCAG TTGTGTCTA TTAAAACAGT TGTGTCAGGG GCACTAGGTA CGTTCTCTCT TTTTGTATGG AAAATGCAAT	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA AGATAATAAG AAAGGAGTTT TCTGTAAAAC TACTAGGGGT CTTAAATGT TTCCATATGT TTCCATATGG ATGGAGGTACAAGA	1111 1211 1211 1411 1511 1611 1711 1811 1911 2011 2111 2211
CTAMATICAL GETTETAAGA GETTETAAGA GETTETAAGA CCCTTEGGGA CCGTTGAGTGA CGGTGAGTGA ATAGGGATAA ATAGGAGAAA CTGAGATAAA CTGAGATAAA CTGAGAGAAA CTAGGAGAAA CATAGGAGAAA CATAGGAGAAA CATAGGAGAAA CATAGGAGAAA CATAGGAGAAA CATAGGAGAAA CATAGGAGAAA	GAAATAACGA GCAAGGCCAT GCCTGAAGGG ACCTGAAGGG TTTGCTAGA TTTGCTAGA TTGTAAAGTG GAACACATTA GGCATTTTAT GTCAGACAGA TTGTAAGGAGA GCATATAGGAG ACATAAGGAG GCATACAGAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAG	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATTGCATT TATGCATCAA CTCTTTGATC AATGGGGAA GCAAAGAAGG TAAACTAGGC TACCACTGAT GACTATTGCA	TAGAGGTT TCATAGGTT TCATAGGTT TAGAGTCCC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAAGCAGCA CAATTGATAG ATCACTCGTG AGACTGGTG	CAGAAAGCTG GACATCTGGG TGGGGAAAGG TGGGGAAAGG TGATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG TGAGGCCAGG AGAAGAATA CAAAGACTCA ATCAGATAAT TGATCCCAG GGAGCCCAG GGAGCCCAAG	ACTGECCTTE AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TTACTGAGTT TGAGAGGTTA GCTTTGCAAG GCAGCACTCT ATTTCACTG GCTTTGCAAG AGGCTAGAAC AGGCTAGAAC	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTTAAGGAG CCGCTAATGAA CTGAACAGTA	TGTCAACTIT TCCAGACATC AAAGAAAATT TTTCTTCTAT TCTGCAGCCA TAAACTCTCA AATTTGAAA TGCCAGGTAC GATAMAATGT GATAGATAGGAG ATTTCAGGAG ATTTCAGGAG ATTTGACAAT	ATCTCAGCTG TGTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGGGCTTCTA TTAAAACAGT TTGTCAGGG GCACTAGGTA COTTCTCTT TTTTGTATGG AAAATGCAAT GGGAGGGG	TGAATCAGA CTCACAAGC AGATGATGAA GAATAATGA AAAGGAGTTT TCTGTAAAAG CTTAAAGGAGT CTTTAAATGT TTCCATATGG ATGGAGGTA AGGAGGTAT CGTACCAGA AGTTCAAATG	2111 1211 1511 1611 1711 1811 1911 2011 2111 2211 2311
CELAMATOLI AGGACTANAA G GETTGTNAGA G CETTGCAGA A CETTGCAGA A CEGTGAGTGA TA ATGGGGATAA A ATAGAGNAG G CTGATATAG G ACTATCACAA G CCAGAGTACAA G ACTATCACAA G ACTATCACA	SAMATANCGA GCAAGÓCCAT GCATCAGGC ACCTGAAGGG AGGGGCCAGG ITTGCTGAGA ITTGTAMAGTG ACCAGACTTC GAACACATTAT GGCATTTTAT GTGAGACAGG MCATAAGGAG GCAGTCCTGA AGTAGAAGAG GTAAGAAGAG	TAMAMAAA TAGATATATC CCCTTGANGA AAGTTANGTA GGAATTTTTC TGTTTGCATC TATGCATCAA CCCTTTGATC AATGGGGAA GCAAAGAAGG TAAACTAGGC TACCACTGAT GACTATTGCA GTATGACTTA	TAGAGGTT TCATAGTGT TCATAGTGT TCATAGTGT TAGAGTCCC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAAGCAGCA CAATTGATAG ATCACTCGTG AGAGCTGGTC GCTGATTTA ATTACTGCG ACCATCTGCG ACCATCTGCG	CAGAAAGCTG AAACTAGCAA GACATCTGGG TGGGGAAAGG TGAGTATAAT TCCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAGATAA TAAGACTCA ATCAGATAA TATGGAAGG TATGGAAGA TATGGAAGG TATGGAAGA TATGGAAGG TATGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG	ACTGGCCCTG AGGACAAGC ATAGATAAGA AAGGAGAAG GCAGCACTCT TTATTATTGG TTACTGGGTT TGAGAGGTTT TTTTCATCTG GCATTGCAAG GAATGGTAACA AGGCTAGAAC GAATGGCTAGAAC	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTAGG ACAACTACGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTAAGGAG CCCTAATGAG CTGAACAGTA GAGCTCTTAGGA	TOTCANCTIT TCCAGACATE ANAGAMATT TTTCTTCTAT ACCTGGCTT TCTGCAGCCA AAATTTGAAA TGCCAGGTAC GATAMATGT GATAGGTAGG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTGGACAAT AGATAGGTAGG GCAGTACAAA GACTTTGGACAAT	ATCTCAGCTG TOTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGGGCTTCTA TTAAAACAGT TGTGTCAGG GCACTAGGTA CGTTCTCTT TTTTGTATGG AAAATGCAAT GGGAGGGGGG TGATGTGTATG	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTTT TCTGTAAAAC CTTAAAGGGT CTTAAAGGGT TTCCATATGG ATGGAGGTAT CGTACAAGG ATGGAGGTAT CATACAGGT ATGGAGGTAT CATACAGGT ATGGAGGTAT CATACAAGGT ATGGAGGTAT TATGGAGGTAT TATGGACCACA TATGGACCACACA TATGGACCACA TATGGACCACACA TATGGACCACACACA TATGGACCACACA TATGGACCACACA TATGCACACACACACACACA TATGCACACACACACACACACACACACACACACACACACA	1111 1211 1411 1511 1611 1711 1811 1911 2011 2211 2311 2411 2511
CELAMATOLI AGGACTANAA G GETTGTNAGA G CETTGCAGA A CETTGCAGA A CEGTGAGTGA TA ATGGGGATAA A ATAGAGNAG G CTGATATAG G ACTATCACAA G CCAGAGTACAA G ACTATCACAA G ACTATCACA	SAMATANCGA GCAAGÓCCAT GCATCAGGC ACCTGAAGGG AGGGGCCAGG ITTGCTGAGA ITTGTAMAGTG ACCAGACTTC GAACACATTAT GGCATTTTAT GTGAGACAGG MCATAAGGAG GCAGTCCTGA AGTAGAAGAG GTAAGAAGAG	TAMAMAAA TAGATATATC CCCTTGANGA AAGTTANGTA GGAATTTTTC TGTTTGCATC TATGCATCAA CCCTTTGATC AATGGGGAA GCAAAGAAGG TAAACTAGGC TACCACTGAT GACTATTGCA GTATGACTTA	TAGAGGTT TCATAGTGT TCATAGTGT TCATAGTGT TAGAGTCCC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAAGCAGCA CAATTGATAG ATCACTCGTG AGAGCTGGTC GCTGATTTA ATTACTGCG ACCATCTGCG ACCATCTGCG	CAGAAAGCTG AAACTAGCAA GACATCTGGG TGGGGAAAGG TGAGTATAAT TCCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAGATAA TAAGACTCA ATCAGATAA TATGGAAGG TATGGAAGA TATGGAAGG TATGGAAGA TATGGAAGG TATGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG	ACTGGCCCTG AGGACAAGC ATAGATAGA AAGGAGAAG GCAGCACTCT TTATTATIGC TTACTGGGTT TGAGAGGTTT TTTTCATCTG GCATTGCAAG GAATGCATA AGGCTAGAAC GAATGCTAGAAC	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTAGG ACAACTAAGGA GGTCTTAAAA AACGTGAGGT ACCTAAGGA CCCTAATGA CTGAACAGTA GAGCTCTTAGGA	TOTCANCTIT TCCAGACATE ANAGAMATT TTTCTTCTAT ACCTGGCTT TCTGCAGCCA AAATTTGAAA TGCCAGGTAC GATAMATGT GATAGGTAGG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTGGACAAT AGATAGGTAGG GCAGTACAAA GACTTTGGACAAT	ATCTCAGCTG TOTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGGGCTTCTA TTAAAACAGT TGTGTCAGG GCACTAGGTA CGTTCTCTT TTTTGTATGG AAAATGCAAT GGGAGGGGGG TGATGTGTATG	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTTT TCTGTAAAAC CTTAAAGGGT CTTAAAGGGT TTCCATATGG ATGGAGGTAT CGTACAAGG ATGGAGGTAT CATACAGGT ATGGAGGTAT CATACAGGT ATGGAGGTAT CATACAAGGT ATGGAGGTAT TATGGAGGTAT TATGGACCACA TATGGACCACACA TATGGACCACA TATGGACCACACA TATGGACCACACACA TATGGACCACACA TATGGACCACACA TATGCACACACACACACACA TATGCACACACACACACACACACACACACACACACACACA	1111 1211 1411 1511 1611 1711 1811 1911 2011 2211 2311 2411 2511
CHAMATOLI AGGACATANA GOTTOTAMA GOTTOTAMA CCCTTGAGGA CCCTTGAGGA ACCACTTAGTT ATGGGATAN ATAGAGANA CCTGATATANG GATTAGACA ACTATAGACA CCAGGACAGA ANATTAGAGA ANATTAGAGA ANATTAGAGA ANATTAGAGA ANATTAGAGA ANATTAGAGA ANATTAGAGA ANATTAGAGA GTCCAACACA GTCCAACACA GTCCAACACA	GAAATAACGA GCAAGGCCAT GCATCTGCC ACCTGAAGGG AGGGGCTAGG TITTGCTAGAA TITTGTAAAGTG GAACACATTA GGCATTTTAT GGCATTTTATA GCATAAGGAG ACTAAAGGAG ACTAAAGGAC TCATTTATAA GCAGAAGAACTG TCATTTATAA	TAMAMAAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATTGCATT TATGCATC TATTGCATC TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC GCAAAGAAGG TAAACTAGGC TACCACTGAT GCCTATTTGCA GTATGACTTA TGGCTCTTTA AAATAAGATG TACAATAACA	TAGAGAGGTT TACAGAGGTT TCATAGATGC TACAGATAGCATAGATAGATAGATAGATAGATAGATAGAT	CAGAAAGCTG AAACTAGGA AAACTAGGA AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAAGACTCA ACAAGACTCA ATCAGATATAT TGCATGCCCAG GGAGACATAA TATGGAAGG CAAGACTCA TATGGAAGG CACAGTCTGA TCTGGTGGGA CTGGCTAGA	ACTGECCTTG AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTATTGAGTT TGAGAGGTTT GTATCATCAG GCTTTGCAAG GCATCATTAA AGGCTAGAAC CAAATGGCTA ATCCTGAAAG CTTTTTTGTAG CTTTTTTTGAAG CTTTTTTTTGAG CTTTTTTTTGAAG CTTTTTTTTAAA	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTCAT CTTCTGCTAT ACACTATCGG CAATATATGT GGTCTTAAAA AACGTGAGGG ACCTAAGGAG CCCTAATGAA CTGAACAGTA GACTCTTAGGA GATCTTTAGGAG GATCTTTAGGAG GATCATTAGAAGTGA GATCAAGGGAG GATCAAGGGAG	TGTCAACTIT TCCAGGCATC AAAGAAAATT TTTCTTCTAT TCTGCAGCCA TAAACTCTCA TAACTCTCA TGCAGGTAC GATAMAATGT GATAGATAGGAG ATTTCAGGAG ATTTCAGGAG ATTTCAGGAG ATTTCAGGAG TGAGATAGTGT GCAGTACAAA GACTTTCTGT TTACCTTGCA GATAAGGAG GATAAGGAT TTACCTTGCA GATAAGGAT TTACCTTGCA	ATCTCAGCTG TGTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGGGCTTCTA TTAAAACAGT TGTGTCAGGG GCACTAGGTA CGTTCTCT TTTTGTATGG AAAATGCAAT TGGAGGGGGC AAAAGAGGGG TTGAGTGAGTAAA AAGCCTTAG CCAAAGTTAC	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTTT TCTGTAAAAGC CTTTAAATGT TTCCATATAGG ATGGAGGTAT TCGTACAAGC ATGGAGGTAT AGTTCAAATGT TATGGAGGTAT TATGGAGGACAC AGTTCAAATG TATGGACCAC AGTTCAAATG TATGGACCAC TGAGCACAGA TGAGCACAGT TGAGCACAGT TGAGCACAGT TGAGCACAGT TGAGCACAGT TGAGCACAGT TGAGCACAGT TGAGCACAGT TGAGCACACT TGAGCACAGT TGAGCACACT TGAGCACT TGAG	1111 1211 1411 1511 1611 1711 1811 1911 2011 2211 2211 2311 2411 2511 2611 2711
CHAMATOLI AGGALTANA G GOTTOTNAGA G COTTOTNAGA G COTTOTNAGA G COCTTOTNAGA G COCTTOTNAGA G ATTOTT ATTOGGATAA A ATAGAGANA G CTATATTAG G ATTATAG G ATTATAGAGA ATTATAGAGA ATTATAGAGA ATATTANAGA ATATTANAGA ATATTANAGA ATATTANAGA ATATTANAGA ATATTANAGA TOTNATTOC GTCCAACAA GTTCTACCTA TOTTCCTACCAACAA GTTCTACCTA	GAAATAACGA GCAAGGCCAT GCATCTGCC ACCTGAAGGG AGGGGCCAGG TITTGCTGAGA TITGTAAAGTG ACCAGACTTC GAACACATTAT GGCATTTTAT GTGAGACAGG MCATAAGGAG GCAGTCCTGA AGTAGAAGTAG GCAGTCTTATAA GCAGAAGCAC TCATTTATAA GCAGAAGCAT TAGGAGAGTA	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATC TATGCATCAAA CCCTTTGATC CACTTTGATC CACTTTGATC TACCACCACA GCAAAGAAGA GCAAAGAAGA GCAATTTCCA GTATGACTTA AAATAAGATC TACCATTATA AAATAAGATC TACCATTATA AAATAAGATC TACCATTATA AAATAAGATC TACCATATACAA TACCATATACAA TTTGGAACTC	TAGAGGTT TACAGAGGT TCATTAGTGT TACAGATGC TACAGATGC TAAGGATAGA TTCATGCTGC TAAAGCAGCA CAATTGATAG ACAGCTGGT GGCTGATTTA ATTATCTGCG ACCATCTGGG GTCTGACTGC TACAGCTGCTATTA ACAGCTGCTTA ACAGCTGCTTA ACAGCTGCTTA ACAGCTGCTTA ACAGCTGCTTA TTACCTATTA	CAGAAAGCTG AAACTAGCAA GACATCTGGG TGGGGAAAGG CATCAGACA TGAGGCCAGG AGAGAGATAAT TCCATTAGG AGAGAGATAAT TCAGATAAT TCAGATAAT TCAGATAAT TATGAAGG CAGGGAGATAAA TATGGAAGG CAGGGTTGA TCTGGGGGAAATAA TATGGAGGG AGAGTTGAAAGG CAGGGTTGAA TCTGGTGGGA CAGGCTTAAA	ACTGGCCCTG AGTGGTAAA AGGACAAGC ATAGATAAGA AAGGAGAAG GCAGCACTCT TTATTATTGG TTAGTGGGTTT TTTTCATCTG GCATTGCAAG GAATGGTAAA AGGCTAGAAC GAATGGTAAAC GAATGGTAAAC GAATGGTAAAG CTTTTGTGAAC	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTAGG CAATATATGT GGTCTTAAAA AACGTGAGGT ACCTAAGGAG CCCTAATGAA CTGAAGGTA GATCATAGGAG TTTAGGTAT GATCAAGGAA GATTAGGAAG GATTAGGAAG AGTTAGGAAG AGTTAGGAAG	TGTCAACTIT TCCAGGCATG AAAGAAAATT TTTCTTCTAT TCTGCAGCCA AAATTTTGAAA TGGCAGGTAC GATAAAATGT GATGAGTAGG ATTTCAGGAG ATTTCAGGAG ATTTCAGGAG ATTTCAGGAG ATTTGAGAAATGT GCAGTACGG CCAGTACAAA GCATTTGTGT TTACCTTGCT TTACCTTGCT TTACCTTGCAAAAGTT TTGGAAAAGTT TTGGAAAAGTT TTGGAAAAGTT TTGGAAAAGTT TTGGAAAAGTT TTGGAAAAGTT TTGGAAAAGTT TTGGAAAAGTT TTGGAAAAGAT TTGGAAAAGAT TTGGAAAAGAT TTGGAAAAGAT TTGGAAAAAGAT TTGGAAAAGAT TTGAAAAGAT TTGGAAAAGAT TTGGAAT TTGGAAAAAT TTGGAAAAAT TTGGAAAAAT TTGGAAT TTGGAAAAAT TTGGAAT TTGGAAAAT TTGGAAT TTGGAAAAT TTGGAAAAT TTGGAAT TTGGAAAAT TTGGAAAAT TTGGAAAAT TTGGAAAAT TTGAAAT TTGAAAT TTGAAAT TTGAAAT TTGAAT TTGAAAAT TTGAAAAT TTGAAAAT T	ATCTCAGCTG TCTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGGGCTCTA TTAAAACAGT TTGTGTCAGG GCACTAGGTA CGTTCTCTT TTTTGTATGC AAAATGCAAT GGGAGGGGG TTGATGTATGA AAGCCCTTAG AAAGTTAC AAATGCAAT AAGCCCTTAG	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA AAAGGAGTTT TCTGTAAAAG CTAAGGGGT CTAAGGGATC CTTAAATGT ATCAGTATGAG ATGGAGGTAT CGTACAAGG ATGGAGGTAT CCTTACAAGG ATGGACCACA CCTCATGAA TGAGGACCAC CCTCATGAA TGAGGACAAGT TCAGGAGAGAGT TCAGGAGAGAGA	1111 1211 1411 1611 1611 1711 1811 2011 2211 2211 2211 2411 2611 2611 2611 26
CTAMATICAL AGGATAMA G GOTTGTAMA A GOTTGTAMA A CCOTTGAGGA A CCOTTGAGGA A ATAGGGATAA A ATAGGGATAA A ATAGAGAMA G CTGATATAAG CHATATAAG ATATTAAGA ATATT	GLAATAACGA GCAAGGCCAT GCCTGAAGGG ACCTGAAGGG TTTGCTTAGA TTTGTAAAGTG ACCAGACTTTA GGCATTTTAT GGCATTTTAT GCATAAGGAG CCATAAGGAG CTAAGAGAGTG CTAAGAGAACTT TCATTTATAA GCAGAAGCAT TCAGTGAGTG	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TATGCATT TATGCATT TATGCATC AATGGGGAAA GCAAAGAAGG TAAACTAGGC TACCACTGAT GACTATTGCA GCTATTGCA TATGGACTTA TGGGCTCTTA TAGGTATAAAAGATG TAGGACTATTGCA TTTGGAACTTA TAGGACTTA TAGGACTA TAGGACTTA TAGGACTA TAGGACTTA TAGGACTA TAGGACTTA TAGGACTA TAGGACTA TAGGACTA TAGGACTA TAGGACTA TAGGAC	TAGAGGGTT TCATAAGGTT TCATAAGGTT TAGAGTCCC TAAGGATAGA TTCATGCTGC AGATGCCCT TAAGGATAGA ACAATGATAG ACAATGATAG ACAATGATAG ACAATGGTC GGCTGATTA ATTACTGCG GCTCAATTA ATTACTGCG TCACACGCC ACAGTGCTA ACACTGCTA ACATGCTA ATTACTCTAGA ACACTGCTA ATTACTCTAGA ACACTGCTA ATTACTTCTAGA	CAGAAAGCTTC AAACTATGCAA GACATCTGGG TGGGGAAAGG TGGGGTATAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AAGAACTCA ATCAGATAAT CAAAGACTCA ATCAGATAAT CGATGCCCAG GGAGACATAA TATTGGATGCCAG CAGAGTCTGA TCTGTGGGA AAACAGTAGT TTTTGTTGGA TTTTTGTTGGA	ACTGECCTTE AGTGATAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG TTAATTATTGC TTACTGAGTT TGAGAGGTTT TGTTCTCAGG GAGTCATTA TCTGCAACA AGGCTAGAAC AGGCTAGAAC ATCCTGAATG CTTTTTTTTGCAAG TATCATGAGACA AGGCTAGAAC ATCCTGAATG CTTTTTTTTGAG TATCATGAATAAC TACAGCAGGT TACAGCAGGT AGGCTAGAACA ATCCTGAATG CTTTTTTTTGAGA TACAGCAGGT TACAGCAGGT AGGCTAGAGACA AGGCTAGACA ATCCTGAATG CTTTTGTGAG TACAGCAGGT AGGCTAGAGGC TACAGCAGGT AGGCTAGAGGC TACAGCAGGT AGGCTAGAGGGGA AGGCTAGAGGCAGAGACA AGGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT ACAGTTTCG ACACTACGG CAATATATGT GGTCTTAAAA ACCGTCAGGA CCGCTAAGGA CCGCTAATGAA CTGAACAGTA AGACCTTCGG TTTAGTATG GATCAAGGGG GTTTGGGAG GTTTGGGGGGGGAT CCGGTAGGGAG	TOTCANCTITI TCCAGGACATE ANAGAMATI TITCTTCTAT RCTGCAGCCA ANACTCTCA ANTITIGAAA TGCAGGTAC GATAMATGT GATAMATGT GATAMATGT GATAMATGT GATAMATGT GATAMATGT GATAMATGT GATAMATGT GATAMATGT GATAGGAGAT ATGATAGGAG ATTGGATAGAAA ATGATAGGAA TTACCTTGCA TTAATGGAT TTACCTTGCA TTGGAGAAGAT CTGGAGAAGAT CTGCAGAAGAT CTGCAGAAGAT CTGCAGAAGAT CTGCAGAAGAT CTGCAGAAGAT CTGCAGAAGAT CTGCAGAT CTGCAGAAGAT CTGCAGAAAAAAT CTGCAGAAAAAAT CTGCAGAAAAAAT CTGCAGAAAAAAT CTGCAGAAAAAAT CTGCAGAAAAAAT CTGCAGAAAAAAAT CTGCAGAAAAAAAAAA	ATCTCAGCTG TGTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGGAGCCAG TTGGGTTCTA TTAAACAGT TTGTTCAGGG GCACTAGGTA CGTTCTCT TTTTGTATGG AAAATGCAAT GCGAGGGGGG AAAAGAGGGG TTGATGTGAT AAGCCCTTAG GCAAAGTTAC AAACCCATAGAT AGCATGGATA AGCATGGATA AGCATGGATA AGCATGGATA AGCATGGATA AGCATGGATA ACAAGCTCCT	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTT TCTGTAAAAC TACTAGGGGT CTAAGGGATC CTTCAAATGT TCCATATGG ATGGAGGTAT TCCATATGG ATGGAGGTAT TATGGAGCACA CCTCTATGAA TGTAGACCAC CCTCTATGAA TCAGGACACAT TCAGGAGACACT TCAGGACACT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACT TCAGGACACT TCAGGACT TCAGGACACT TCAGGACT T	1111 1211 1411 1511 1611 1711 1811 2011 2211 2211 2311 2411 2511 2611 2711 2011 2011
CHAMATOLI AGGACTANAA G GETTGTAGAA A CCCTTGAGGA A CCCTTGAGGA A CCGTGAGTGA TA ATGGGATAA A ATAGGAGAAG ATTAGGAGAAG ATTAGGAGAAG ATTAGGAGAAG ATTATAGGA ATTAGAGAAAG CTAGATTAAA CTAGGAGAGA ATTATAGGA ATTATATAGGA ATTATATAGGA ATTATATAGGA ATTATATAGGA ATTATATAGGA ATTATATAGGA ATTATATAGGA ATTATATAGT TCTGAACAA TTATATTAGT TTATATTAGT TTATATTAGT TTATATAGT TTATATTAGT TTATATAGT TTATATAGT TTATATAGT TTATATAGT TTATATAGT TTATATAGT TTATATAGT TTATATAGT TTATATAGT TTATATATA	GAAATAACGA GCAAGGCCAT GCATCTGCC ACCTGAAGGG AGGGGCTAGG TITTGCTGAGA FITTGCTAAGGT GCACTCTTAT GGCATTTTAT GGCATTTTAT GCATCCTGG ACACAAGGAG GCAGAGGAG TCATTTATAA GCAGAAGCAT TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA	TAMAMAAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATTGCATT TATGCATC TATTGCATC TATTGCATC TATAGCATC TATAGCATC GCAAAGAAGG GCAAAGAAGG TAAACTAGGC TACCACTGAT GACTATTGCA GTATGCATTA TGGCTCTTTA AAATAAGATG TTTGGAACTC GTGGTTGTAT TTTGGAACTC GTGGTTGTAT	TAGAGGTT TACAGGGTT TCATAGGTT TACAGATOC TAGGATCCC TAAGGATAGA TTCATGCTGC TAAGGATAGA ACATGCTGT AAAGCAGGA ATCACTCGTG AGAGCTGGTTA AGTACTCGTG ACCATTCGGG ACCATCTGGG ACCATCTGGC ACCATCTGGT TAGCTATTA	CAGAAAGCTE AAACTAGCAA GACATCTGGG TGGGGAAAGG TGGGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAAGAATA CAAAGACTA ATCAGATATAA TGATACTCA ATCAGATAAT TATGAAGACTA TATGAAGATAA TATGAAGGC GGAGACATAA TATGAAGGC CAGAGTCTGA TCTGGTGGGA TTTTTTTTGAC TTTTTTTTGAC TTTTTTTTGAC TTTTTTTTGAC TTTTTTTTGAC TTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTGAC TTTTTTTTTT	ACTGECCTTE AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TTTCATCTG GCTTTGCAAG GCAGCACATA AGGCTAGAAC GAAATGGATA ATCCTGGAAG CTATTAGAAG CTATTAGAAG CTATTAGAAT CTTTTGAAG CTATTAGAAT CTAGAATAG TATGAATAAC TACAGCACGT AGGGTCAGGA	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT ACACTATCGG CAATATATGT GGTCTTAAAA ACGTCAAGGA CCGTAATGAA CTGAACAGTA GAGTCTTGGG GATTAGGAG	TOTCHACTITE TCCAGACATE ANAGAMATIT TTCTTCTAT ACTIGGETAT TCTGCAGCCA ANTITTGAAA TGCCAGGTAC GATAMATGT GATAGATAGG ATTGGACAA ATTGGACAAT AGATAGTAGG CCAGTACAAA CACTTTGGAT TTACCTTGCA TTACCTTGCA TTACTGGAT TTACCTTGCA TTACTGGAT TTGGAAAGAT TTGGAAAGAT TTGGAAAGAT CTGGAGAGT CTGGAGAGT CTGGGAGAGT CTGGGAGAGT CTGGGAGAGT CTGGGAGAGT CTGGGAAAGAT CTGGAGAGT CTGGGAGAGT CTGGGAGAGT CTGGGGAGAGT CTGGGGAGAGT CTGGGGAAGAT CTGGGAGAGT CTGGGGAAGAT CTGGGAGAGT CTGGGGAAGAT CTGGGGAAGAT CTGGGAGAGT CTGGGGAAGAT CTGGGGAAGAT CTGGGGAAGAT CTGGGAGAGT CTGGGGAAGAT CTGGGGAAGAT CTGGGGAAGAT CTGGGGAAGAT CTGGGGAAGAT CTGGGGAAGAT CTGGGGAAGAT CTGGGAGAGT CTGGGGAAGAT CTGGGGAAGAT CTGGGAGAGT CTGGGAAAA CCCTACCCCC CTGCCCCC CTGCCCCCC CTGCCCCCC CTGCCCCCC CTGCCCCCC CTGCCCCCC CTGCCCCCC CTGCCCCCC CTGCCCCCC CTGCCCCCCC CTGCCCCCCC CTGCCCCCCC CTGCCCCCCCC	ATCHAGGEGG TGTTGGAMA ATANGTGGTG TGGGGCTGGG TTGGGCTTCTA TTAMACAGT TTGTGTCAGGG GCACTAGGTA CGTTCTCTCT TTTTGTATGG AMATGCAGT CGGAGGGGG AMAGGGGGGG TGAGTGGTT AGCATGGATA AGCCTTAG AGCATGGATA AGCATGGATA CAGGGCTGA AGCACTTAGATA CAGGGCTGATA CAGGCTGATA CAGGGCTGATA CAGGGCTGATA CAGGGCTGATA CAGGGCTGATA CAGGCCTGATA CAGGCCT	TGAATCAGA CTCACAAGC AGATGATGAA GAATAATGA AAAGGAGTTT TCTGTAAAAG CTTAAATGG TTCCATATGG ATGGAGGTAT CGTACCAAGA AGGTATT CGTACCAAGA AGGTACT TATGGACCAC CTTTAAATG TATGGACCAC TCAGGAGACAG TCAGGAGACAG TCAGGAGACAG TCAGGAGACAG TCAGGAGACAG TCAGGAGACAG TCAGGAGACAG TGTGTGATACC AGGGCATAAA	1111 1211 1411 1611 1611 1771 1911 2011 2011 2211 2211 2211 2211 221
CHAMATOLI AGGALTANA G GOTTOTNAGA G COTTOTNAGA G COTTOTNAGA G COCTTAGGA A COCTTAGGA A COCTTAGGA A ATAGGARAA G ATAGGARAA G ACTATGACA G ACTATGACA G ACTATGACA G ACTATGACA G ATATTAAGGA AIATTAAGGA AIATTAAGGA AIATTAAGGA AIATTAAGGA AIATTAAGGA AIATTAAGGA AIATTATGC TOTNATTCC TOTNATTCC TOTNATTCTC TACTAGCA TACTACAT TACTACATTCTA TACTACATGA TACTACATAGGA TACTACATAGA TAC	GLAATAACGA GCAAGGCCAT GCCTGAAGGG ACCTGAAGGG TTTGCTTAGA TTTGCTTAGA TTTTAAAGTG GCACTTTATA GCACTTTATA GCACTTCTGA GCACTCCTGA ACATGAAGTG GCACTCCTGA ACATGAAGTG TTAGTAGAAGTG TTAGTAGAAGTG TTAGTAGAAGTA TAGTTGAGTAACA TGCTGGAGTA TGCCCCGAGA	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATTGCATT TATGCATC TATTGCATC TATGCATCAAC CCCATTGATC AATGGGGAAA GCAAAGAAGG TAACACTGAT GACTATTGCA GTATGACTTA TGGACTTTA TTTGGAACTC GTGGTGTTAT AGTGAGGACC CTGGTGTGTAT AGTGAGGACC CTGGTGGTGTTAT AGTGAGGACC CCTGGAGGACC CCTGGAGGACC CCTGGAGGACC CCTGGAGGACC CCTGGAGGACC CCTGGAGGACC CCTGGAGGACC CCTGGAGGACC CCTGGAGGACC	TAGAGGGTT TCATTAGGGT CTTCAGATGC TAGGATCCCC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAAGCAGCA CAATTGATAG ATCATCCGTG GGCTGATTTA ATTATCTGCG GTCTGACTGG GTCTGACTGG TAAAGCAGTATTA ATTATCTGCG TACATCTGGG TAGCTGACTGG TAGCTGATTA ATTATCTTCTGT TAAATAT	CAGAAAGCTE AAACTAGCA GACATCTGGG TGGGGAAAGG TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAATAA CAAAGACTCA GGAGCACATAA CAGAGACATAA TATGGAAGGG CAGAGTCTGA CAGGCTACAT AAACAGTAGT TTTTGTGGC TGGCCTACAT TGGACATTA	ACTGECCTTG AGTGATAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG TTAATTATTGC TTACTGAGTT TCACGAGTT TCACGAGTT TCTCACG GCTTTCCAAG GCACACCACA TCCCGCAACA AGGCTAGAAC TCTCGCAACA ATCCTGAATG TATGATAAC TACAGCAGG TATGATAAC TACAGCAGG ATGCCTCAGG	TGGTTCCCAC TGGTGAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT ACAGTTTGG ACAACTACGG CAATAATGT ACCTAAGGAG ACCTAAGGA CCCTAACGAGT ACCTAAGGAG CCTAACGAGT TTTAGTATG GATCTAGGAG TTTTAGTATG AGTTCAGGAGT AGTTAGGAGT AGTTAGGAGT AGTTAGGAGT AGGTTAGGAACAGT TGGGCAT TGGGTTCAGCT AGACCATAGA TGGTTCAGCT AGACCATAGA TGGTTCACTC AGACCATAGA TGGTTCACTC AGACCATAGA TGGTTCACTC AGACCATAGA TGGTTCACTC AGACCATAGA	TOTCANCTIT TCCAGGACTA ANAGAMANT TITTCTTCTAT ACCTGGCTTT RCTGCAGCCA ANATTTGAMA TGCCAGGTAC GATAMANTGT GATGAGTAGG ATTTCAGGAG ATTTCAGGAG CAGTACAA AGGATACAA AGGATACAA AGGATACAA AGGATACAA AGGATACAA CACTTTGTGT TTACCTTGCA AGTAGGAGAT TTGGAAAGAT TTGCCAGC	ATCHAGETG TGTTGGAAA ATAAGTCGTG TGGAGCCTGG TTGTAGCCAG TTGAGCTAGTA TTAAAACAGT TGTGTCAGGG CCACTAGGTA AAAATGCAAT CGGAGGCGCC AAAAGAGGGG TTGATGTGA AAGCCCTTAG CCAAAGTTAC AAAATGAAAT AAAATGAAAT AAAAGAGGGG CCAAAGTTAC AAAATGAAAT AAAAGAGGCTGC CCAAAGTTAC AAAATGAAAT AAAAGCCTGTAAAAT AAAAGCCTGTAAAAT AAAAGCCTGTAAAAT AAAAGCCTGTAAAAT	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTTT TCTGTAAAAC TACTAAGGGT CTAAGGGATC CTTAAAATGT TCCATATGG ATGGAGTAT TCGGTAATGA ATGAGTAT TCAGAGTAT TCAGAGTAT TCAGAGTAT TATGGAGTAT TATGGACTACA CCTCTATGAA TCAGCACAGT TCAGGAGACAG TCAGGAGACAG TCAGGAGACAG TCAGGAGACAG TCAGGAGATAAA ATATATAAAGG	1111 12111 1411 15111 1611 17711 1911 2011 22111 22111 22111 22111 22111 2311 2311 2311 2311 2311 2311
CHAMATOLI  GOTTOTAMA G GOTTOTAMA A  GOTTOTAMA A  COTTCAGGA  COTTCAGGA  ACCOTTGAGTA  ATGGGATAA  ATAGAGANA  CTATATAG  ATTAGAGAA  ATATACACA  GATTATAG  ATTAGAGAA  ATATAAGAA  ATATAATTAC  TOTAACTTC  TOTAACTTC  TATACTTCAACAA  TATACTTCAACAA  ACAAGAATAAA  ACAAGAATAAA  ACAAGAATAAA  ACAAGAATAAAA  ACAAGAATAAAA  CCAAGAATAAAA	GAAATAACGA GCAAGGCCAT GCATCTGCC ACCTGAAGGG AGGGGCCAGG TITTGCTAAAGTG ACCAGACTCC GCATTTTAT GGCATTTTAT GGCATTTTAT GTAAGAGAG GCAGAGAAC TTAAGAGAAC TTAAGAGAAC TTAAGAGAAC TTAAGAGAAC TTAAGAGAAC TTAGGTGAGAAC TTAGGTGAGAAC TGCCCGGAGA AAAAGCTGAG GAGAAAGCG	TAMAMAAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TOTTTGCATT TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC TACACTGATC GACTATTGCA GCAAAGAAGG GTAAACTAGGC TACACTGAT GACTATTGCA GTATGCATCTTA AAATAAGATC TACACTATACA TTTGGACTTTA AAATAAGATC GTGGTTGTAT TACGTTGTATACA GTGGGACCC GGGAACCAGA AAATATTCT	TAGAGAGGTT TACAGAGGTT TCATAGGTT TAGAGTCCC TAAGGATAGC TTAGATCCT TAAGGATAGA TCATGCTGT TAAAGCAGCT CAATTGATAG ATGATAG ATGATAG ATGATAG ATGATAGATA	CAGAAAGCTG AAACTAGGAA GACATCTGGG TGGGGAAAGG CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAAGAATA CAAAGACTCA ATCAGATATAAT TGCATGCCCAG GGAGACATAA TATGGAAGGC CAGAGTCTGA TCTCGTGGGA TCTCGTGGGA TTTTGTTGAC TGTGCCCTAC TGGACATTTA TATTATAAAA CCTCATTTTTTA	ACTGECCTTE AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTACTGAGTT TTTCATCTG GCTTTGCAAG GCACCACAA ACGCCACAA ACGCTAGAAC CAATTGCAAG CTTTTGTGAG TATGATTGCAAG CTTTTGTGAG CTTTTGTGAG CTTTTGTGAG ATCCTTGTTGAAG CTTTTGTGAG ATCCTTGTTGAG ATCCTTGTTAGAG AGGGTCAGGG AGGGTCAGGG AGGGTCAGGG AGGGTCAGGG AGGGTCATGG	TGGTTCCCAC GTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT ACACTATCG ACACTATCG GGTCTTAAAA AACGTGAGGT ACCTAAGGAG CCGCTAATGAA CTGAACAGTA FAGTCTTGGG GATTAGGAG GATTAGGAG GATTAGGAG GATTAGGAG GTTTGGGAG GTTTGGGAG GTTTGGGAG GTTTGGGAG GTTTGGGAG GTTTGGGAG CCGGACTAGA TGGTTCACTC AGAGGAACAG	TOTCANCTIT TCCAGGARAGA ANAGAMANTI TITCTTCTAN ACCTGGCTT TCTGCAGCCA ANATTTGANA TGCCAGGTAC GATANANTGT GATGAGTAGG ATTCAGGGA ATTCAGGAGAG ATTGGACAAA AGGATAGGG CCAGTACAAA GACTTTGTGT TTACCTTGCA GATANAGGAT TTGGAAAAGT TTGGAAAGAT TTGGAAAGAT TTGGAAAGAT CTGGAGAGTA CTGGGAGAG CTGTGGTAAG AGATTAACCA GAGTTAATCA AGATCAACAA AGATCAACAA AGATCAACAA AGATCAACAA AGATCAACAA	ATCTCAGCTG TOTTGGAMA ATAMOTOGTG TOGAGCCTGG TTGGGCTTCTA TTAMACAGT TTGTGTCAGG GCACTAGGTA CGTTCTCTCT TTHIGTATGG AMATGCAGT CGGAGGGGG AMAGGGGGG TTGAGTGGT AMGCCTTAG CCAAGGTGAT CCAAGGTGAT CCAAGGCCTTAG AMTCTGAAT CAAGGCCTGT ACAAGTTAC AMTCTGAAT CAAGGCCTGT TAGAAGTG TTAGAAGTG TTAGAAGTG TTAGAAGTG	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTTT TCTGTAAAAG CTAAAGGAGTT TCCATATAGG ATGGAGGTC CTTAAATGT ATGGAGGTAT CGTACCAAGA TGTACCAAGA TATGGACCAC CCTCTATGAA TCAGGACCAC TATGGACCAC TATGGACCAC TATGGACCAC TATGGACCAC TATGGACCAC TATGGACCAC TCAGGACACGT TTAGGACACC AGGGCATAAA ATATATAAGG TGGATGGATA TTTAGACCAG	1111 1211 1411 1511 1611 1771 1811 2011 2211 2211 2211 2511 2611 2911 2911 3011 3111 3111
CTAMATICAT  GOTTOTRAGA  GOTTOTRAGA  CCOTTGAGGA  CCOTTGAGGA  ACCOTTGAGGA  ATAGAGRANG  ACTATGAGA  ACTATGAGAA  ACTATGACAA  ACTATGACAA  ACTATGACAA  ACTATGACAA  ACTATGACAA  ACTATGACAA  ACTATGACAA  ACTATGACAA  ACTATGACAA  ACTATCACAC  ACTATCACAC  GOTCAACACA  GOTCAACACA  ACTATCATTOTA  TACTACTTOTA  TACAACAGGA  ACAGGAGTTA  ACAGGAGTTA  ACAGGAGTTA  ACAGGAGTTA  ACAGGAGTTA  ACAGGAGAGTA  ACAGGAGAGTA  ACAGGAGATA  ACAGGAGATA	GLAATAACGA GCAAGGCCAT GCCTGAAGGG ACCTGAAGGG TTTGCTTAGA TTTGTAAAGTG ACCAGACTCTA GGAACACATTA GGCATTTTAT GGCATTATAA ACCAGAAGGAG CTAGAAGGAG CTAGAAGGAG TCATTATAAA GCAGAAGCATA TGCCCCGAGA ACCAGAAGCATA TGCCCCGAGA AAAAGCTGAG AAAAGCTGAG CATAAACTCA	TAMAMAMA TAGATATATC TAMAMAMA TAGATATATC CCCTTGAMGA AMGTTANGTA GGANTITTC TATGCATT TATGCATT TATGCATC TATGCATC TACCACTGAT GACACTGAT GACACTGAT TACCACTGAT TACC	TAGAGGGTT TCATAAGGTT TCATAAGGTT TAGAGTCCC TAAGGATAGA TTCATGCTGC AGATGTCCTT TAAGGATAGA ATCATCGTG AGAGGGTCATTA ATTATTCTCGTG GGCTGATTTA ATTATTCTGGG GTCTGACTGG ACAGTGCTA ATTATTTTCTGTG ACAGTGCTA ATTATTTTTTTTGTATAATAT TGCTGAGAAACC ATAGAGAGAACC ATAGAGAGAACT TTAGATCATTTAAATAT TGGGAGAACC ATAGAGTGAT TTAGATCATTT	CAGAAAGCTTG GACATCTGGG TGGGGAAAGG TGGGGAAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TCAGGCCAGG AGAGAATAA CAAAGACTCA ATCAGATAAT CAAAGACTCA ATCAGATAAT CAAAGACTCA ATCAGATAAT TCATGCCAG GGAGACATAA TCTGTGGGA CTGGTGATTGAG TTTTGTTGAC TTTTGTTTGAC TGGACATTTAAAA CCTGATTGTG CATGATTGTG CATGATTGTG CATGATTGTG CATGATTGTG CATGATTGTG CATGATTGTG CAAATTGTAAAA	ACTGECCTTE AGTGATAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG THATTATTGC TTACTGAGTT TITTCTCTGAGTT TITTCTCTGAGTT TITTCTCTGAGTT TCTGCAACA AGGCTAGAAC ATCCTGAATA AGGACTACGCA ATGCCTTCTT AGGGCCATGGA ATGCCTTCTT ACAATAGCCA AGAATAGCCA AAGTAGGCCC ACAATAGCCA AAGTAGGCCC ACAATAGCCA AAGTAGGCCC ACAATAGCCA AAGTAGGCCC ACAATAGCCA AAGTAGGCCC ACAATAGCCA AAGTAGGCCC ACAATAGCCA AAGTAGGCCC AAGAATAGCCA AAGTAGGCCC AATAGCCA AAGTAGCCCA AAGTAGGCCC ACAATAGCCA AAGTAGGCCC ACAATAGCCA AAGTAGGCCC ACAATAGCCA AAGTAGGCCC AATAGCCA AAGTAGGCCC AAGAATAGCCA AAGTAGGCCC AAGAATAGCCA AAGTAGGCCC AAGAATAGCCA AAGTAGGCCC AAGAATAGCCA AAGTAGGCCC AAGAATAGCCA AAGTAGGCCA AAGTAGGCCA AAGTAGGCCA AAGTAGGCCA AAGTAGCCA AAGTAGGCCA AAGTAGGCA AAGTAGAATAGCA AAGTAGGCA AAGTAGGCA AAGTAGGCA AAGTAGGCA AAGTAGGCA AAGTAGAATAGCA AAGTAGGCA AAGTAGGCA AAGTAGGCA AAGTAGGCA AAGTAGGCA AAGTAGAATAGCA AAGTAGAATAGCA AAGTAGGCA AAGTAGAATAGCA AAGTAGGCA AAGTAGAATAGCA AAGTAGGCA AAGTAGAATAGCA AAGTAGAATAGCA AAGTAGGCA AAGTAGGAATAGCA AAGTAGAATAGCA AAGTAGAATAGCA AAGTAGAATAGCA AAGTAGAATAGCA AAGTAGAATAGCA AAGTAGAATAGCA AAGTAGAATAGAA	TGGTTCCCAC TGGTGAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT ACAGTTTGG ACAACTACGG CAATATATGT ACCTTAAGAA ACCGTCAGGT ACCTAAGGAG CGCTAATGAA TTTTAGTATG GATCATGGG TTTTAGTATG TGTTTGGGGAT TGTTGGGAT TGGTTCAGCAT TGGTTCACTC AGAGGAACAGT AGTTCACTC AGAGGAACAGT AGTTCACTTAACAGT AGATTCACTC AGAGGAACAGT AGATTCACTC AGAGGAACAGT AAAAGGGAG AAAAAGGGAG AAAAAGGGAG TTTCATTATAA	TOTCANCTIT TCCAGACATE ANAGAMANT TITICTTCTAT RETGCAGCEA ANATITICATAT RETGCAGCIA ANATITICANA TGCCAGGTAC GATAMANTGT GATAMATGT ATTICAGAGA ATTICAGAGA ATTICAGAGAT ATGATAGTAG CACTITICAG GATAMATGT TACCITICA GATAMATGT TTACCTTICA GATAMAGGAT TTGGAGAGT TGGAGAGT TGGAGAGT TGGAGAGT TGGAGAGT TGGAGAGT TGGAGAGT TAGAGAGAG	ATCTEAGETG TOTTGGANA ATANGTCGTG TGGAGCCTGG TTGGAGCCAG TTGGGTTCTA TTANACAGT TTGTTCAGGG GCACTAGGTA CGTTCTCT TTTTGTATGG ANAGAGGGG ANANGAGGGG TTGAGTTGAT AGCCCTTAG ANGCCCTTAG ANGCCCTTAG CCANAGTTAG ANTCGANAT ACCATGGATA ACCATGGATA ACCATGGATA ACCATGGATA ATCANGTGC CCTAGCAAGT ATTANAGTAG TTAGAAGTATA TAGGAAGGA TTAGAAGTATA TAGGAAGGA	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTTT TCTGTAAAAC TACTAGGGGT CTAAGGGATC CTATAGAATGA TCGATCAATGA ATGGAGTAT TCGATCAAGA ATGTCAATATGG ATGGAGCAAGA AGTTCAATGA CCTCTATGAA CCTCTATGAA TCAGGACACAT TCAGGAGCAC CTTTAGGACAC CTTTATGAA TCAGGACACAT TCAGGAGAC TTGTGATACC AGGGCATAAA ATATATAAGG TGGATGGATAA ATATATAAGG AAAGAGCTAC AAAGAGCTAC	1111 1211 1411 1511 1611 1771 1911 2011 2211 2211 2211 2311 2411 2511 2611 2711 2911 3011 3111 3211 3211
CHAMATOLI  GETTETHAGA G GETTETHAGA G CETTGAGGA A CCCTTGAGGA A CCCTTGAGGA A ATAGAGATAA A ATAGAGATAA G CTGATTATAG G ACTATTAGGA A ATAGAGATAA A CCAAGTATAA A CCAAGTATAA A CTAGAGAGAA G ATATTAAGGA A ATATTAAGGA A ATATTAAGGA A ATATTAAGGA A TATTATTAC T GTCCAACAA G TATCTGACTA T TATCTTGA T TATATTTTA T TAAAAAGGA A ACAGGAGTTA T CCAGGAGTTA A CCAGGAGTTA C ACAGGAGTTA C ACAGGAGTTA C ACAGGAGTTA T ACAGGAGTTA T TGTACTACT C	GAAATAACGA GCAAGGCCAT GCAAGGCCAG ACCTGAAGGG AGGGGCCAGG FTTGGTAAAGGG ACCAGACTTCA GCAATTTATAA GGCATTTTATAA AGTAGAAGGAC CTAAGAGAAC CTAAGAGAAC CTAAGAGAAC CTAAGAGAAC CTAAGAGAAC CTAAGAGAAC TAAGAGAAC TAAGAGAAC TAAGAGAAC TAAGAGAAC TAAGAGAAC TAAGAGAAC TAAGAGAAC TAAGAGAC TAAGAGAC TAAGAGAC TAAGAGAC TAAAGAC TTAAAGAGC TTAAAGAG TTAAAGAGC TTAAAGAG TTAAAAGAG TTAAAGAG TTAAAAGAG TTAAAAGAG TT	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATTGCATT TATGCATC TATTGCATC TATGCATC TATGCATC TATGCATC TACCCTTGATC GCAAAGAAGG TAAACTAGGC TACCACTGAT GCCTCTTTA AAATAAGATG TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACCACTGAT TACGACTTA AAATAAGAATT CTGGAACCA CCTGGAGCCC CCTGGAGCCC CCTGGACCAGA AAATTATTCT AAAGGAACTT AAAGGAATT	TAGAGAGGTT TACAGAGGTT TACAGATGC TACAGATAGCAT TACAGATAGCAT TACAGATAGCAT TACAGCATGC TAAAGCAGCA ACACTGCTC ACAGCTGGTC ACACTGGTG ACCACTGGG ACCACTGGGC ACCACTGGG ACCACTGGGTATA ATTACTGGG ACCACTGCTATA ATTACTGGG ACCACTGCTATA ATTGTTCTGT GGGGGAAACCC ATAGAGTGAT TTAGGTGAT GGAGAAATTTA GGGAGAATTTA GGAAAATTTA GGAAAATTTA GATAGATTATT	CAGAAAGCTG GACATCTGGG TGGGGAAAGG TGGGGAAAGA TGCCTTTAGG CATTCAGACA TGAGGCCAGG TGAGGCCAGG GCAGACATAA TATGGATATAAT TGCATTCAGACA TATGGATATAA TATGGAAGATAA TATGGAAGG GCAGACATAA TATGGAAGG CAGAGTCTGA TCTGGTGGGA AACAGTAGT TTTTTTTTGA TGTGCCTAC TGGACATTAA TATTATAAAC TCTTATTTTTTTGA	ACTGGCCTTG AGTGGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGG TTACTGAGTT TGAGAGGTTT GAGAGGTTT GAGAGGTTATATTGCA GCTTTGCAAG GCTTTGCAAG AGGCTAGAAC CAAATGGCTA ATCCTGAATG CTTTTTGTAGG CTTTTTGTAGG CTTTTTGTAGG ATGCTCAGATC TACAGCACGT AGGGTCAGGG AGGGTCAGGG AGGGTCAGGG AGGGTCAGGG AGGGTCAGGG AGGGTCAGGA AGGTAGTAGAC AAGTTAGTCA AAGTAGGGCC AAGTAGGCC	TGGTTCCCAC TGGTGAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTGG ACAACTACGG CAATATATGT GGTCTTAAAA ACGTGAGGT ACCTAAGGAG CCGCTAATGAA CTGAACAGTA GAGTCTTGGG TTTTAGTATG GATCATGGGA GTGTAGGAG GTGTAGGAG GTGTAGGAACAGT AGATCAAGGAA GATAAGGAA GGTTCGGCAT CCGGACTAGA TGGTTCACTC AGAGGAACAG AAAAAGGGAG AAAAAGGGAACAG AAAAAGGGAACAG AAAAAGGGAACAG AAAAAGGGAACAG AAAAAGGGAACAG AAAAAGGGAACAG AAAAAGGGAACAG CATATTATAA	TOTCAGETT TCCAGGAAAATT TCCAGGACAA AAGAAAATT ACCTGGCTT TCTGCAGCCA AATTTGAAA TGCCAGGTAC GATAAAATGT GATAGATAGGAAA ATTGGACAA ATTGGACAA ATTGGACAAT TACCTIGCA TACCTICCAGC TACCTICC	ATCTCAGCTG TGTTGGAAA ATAAGTGGTG TGGGGCTGTG TTGAGGCAG TTGGCTTCTA TTAAAACAGT TGTGTCAGGG GCACTAGGTA CGTTCTCT TTTTGTATGG AAAATGCAGT TGGAGGGGGC AAAAGAGGGGG CAAAGTTAC CAAGCTTAG CAAAGTTAC CAAGGCCTGTA AACCAGTAGAT ACCATGGATA ATCAAAT ACCATGGATG ATTCAAGTGC TTAGGAAGTA ATTCAAGTGC TTAGGAAGTA ATTCAAGTGC TTAGGAAGGA AAAAAGAAGAGA AAAAAGAAGAGA AAAAAGAAG	TGAATCAGA CTCACAAGC AGATGATGAA GAATAATGA AAAGGAGTTT TCTGTAAAAGC CTTTAAATGT TTCCATATGG ATGGAGGTT TCTGACAGC ATGGAGGTAT ATTGACCAC AGTTCAAATG TTCAGAGCACA AGTTCAAATG TTCAGAGCACA TCAGGACACA TCAGGACACA TCAGGACACA TCAGGACACA TCAGGACACA TTATAAACG TGGATGATAAA ATATATAAGG TGGATGGATA AAACAGCTAC AAACAGCTACC AACCACCACA	1111 1211 1411 1511 1611 1771 1811 2011 2211 2211 2211 2511 2611 2911 2911 3011 3111 3111
GETTETRAGA G GETTETRAGA G GETTETRAGA G GETTETRAGA G CCTTGGGA A CCTTGGGA A CCTTGGGA A CGGTGGTGA A ARAGGGAAG G ATATGACA G CGAGTATAG G ATATGACA G ATATGACA G ATATGACA G TOTATCACA G ACAGAAGTA G ACAGAAGTA G ACAGAAGTA G TOTATCACA TOTATCACA G TOTATCACA TOTATCACA G TOTATCACA TOTATCACACA TOTATCACA TOTATCACA TOTATCACACA TOTATCACACA TOTATCACACA TOTAT	GLAATAACGA GCAAGGCCAT GCCTGAAGGG ACCTGAAGGG TTTGCTTAGA TTTGCTTAGA TTTGATAAGTG GCACTTTA GCAATTTAT GCATTTATA GCAGTACGA GCAGTCCTGA ACATGAAGTG GCAGTGAGTA TAGTGAAGTG TAGTTAAAA CCAGAAGCAT TAGTGGAGA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGGTGAGTA TAGATGAGAGCT TAAAACTCA TTAAAACTCA TTAAAAGGCT GATCCTGCCA	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATTGCATT TATGCATC TATTGCATC TATGCATCA CCCATTGATC AATGGGGAAA GCAAAGAAGG TAACACTGAT TAGCACTTTA TAGCACTTTA TAGCACTTTTA TAGCACTTTTA TAGCACTTTA TAGCACTTTTA TAGCATTTA TAGCATTTA TAGCATTTA TAGGACTT TAGGACTT TAGGACTT TAGGACTT TAGGACTT TAGGACTT TAGGACT TAGGA	TAGAGGGTT TACAGAGGTT TCATTAGGTT CTTCAGATGC TAGGATCCCC TAAGGATAGA TTCATGCTGC AGATGCTCC TAAAGCAGCA CAATTGATAG ATCATCCGTG GGCTGATTTA ATTATCTGCG GTCTGACTGG TAAACATCGTG TAAACATCGTG TAAACATCGTG TAGCTATTA ATTATCTGCG TCTCACTGG TTAAATAT TCTTTAAATAT TCTTAAATAT TGGGAGAACC ATAGAGTGAT TTAAGTCATT TAAGATTTA GGTAGAATTTA GAGAATTTA GAGAATTTA GAGAATTTA GGTAGATATA	CAGAAAGCTTE AAACTATGCAA GACATCTGGG TGGGGAAAGG TGGGGAAAGG AGATATAAT TGCCTTTAGG CATTCAGACA TGAGGCAGG AGAGAATAA CAAAGACTCAA ACAGATACTCAG GGAGCATAA TATGGAAGGG CAGAGTACTGAA AAACAGTAGT TTTTTTGGC TGGCCTACAT TATTTTGAC TGGACATTTA TATTTTAAAA CCTGATTGTAG CATTTATAAAA CTTATTTAAAA CTTATTTAAAA CTTATTTAAAA CTTATTTAAAA	ACTGECCTTE AGTGGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAG TTAATTATTGC TTACTGAGTT TCTGGAGTT TCTGGAGTT TCTGCAACA TCTGCCAACA TCTGGCAACA AGGCTAGAAC TATGATATAC TACTGGAATG TATGATATAC TACTGGAATG TATGATATAC TACTGGAATG TACTGGAATG TACTGCAACA AGGCTAGAC TACTGCAACA AGGCTAGAC TACTGCAACA AGGCTAGAC TACTGCAACA AGGCTAGGC AGACTTCCT ACAATAGCCA AAGTGGCC TACTGTTAAA	TGGTTCCCAC TGTCAAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT ACAGTTTCG ACAACTACGG CAATAATGT ACCTCAGGAT ACCTCAGGAT ACCTCAGGAT ACCTCAGGAT ACCTCAGGAT ACCTCAGGAT ACCTCAGGAT ACCTCAGGAT ACCTCAGGAT TTTAGTATG AGTTCAGGAG AGTTAGGGAG AGTTAGGGAG AGTTAGGGAG TTTGGGCAT AGGTCACTAGA TGGTTCACTC AGAGGAACAG AAAAAGGGAG AAAAAGGGAC CCTAATTTAC CCTAATTTAC	TOTCANCTIT TCCAGGACAT ANAGAMANT TITCTTCTAT ACCTGGCTT RCTGCAGCCA ANATITGANA TGCCAGGTAC GATANANTGT GATGAGTAGG ATTTCAGAGA ATTTCAGAGAT AGATAGTAG CCAGTACAAA GCATTACAT TACCITGCA ATTAGANAGAT TTACCITGCA ATTAGANAGAT TCGGAGAGT CTGGAGAGT TCGGAGAGT TAGATCAACAA AGATCAACAA AAATGTAAT AAATGTAAT TAAATGTAAT TAAATGTAAT TCGAGAGAGT TCGAGAGGAG ATTTGACAAA	ATGTEAGGTG TGTTGGAAA ATAAGTGGTG TGGGGTTGTA TTAAAACAGT TGTGTCAGGG CCACTAGGTA TGTGTCAGGG AAAATGCAT TGTGTCAGG AAAATGCAT TGTGTCAGG AAAATGCAT TGTGTATGG AAAATGCAT AAGAGGGG CCAAAGTTAC AATCTGAAAT AGCACTTGA AATCTGAAAT AGCATGAAAT AGCATGAAAT ATCAAGTGC TTAAGGAAGGT TTAAGGAAGGA AAAAGAAGGT TTAAGGAAGG	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTTT TCTGTAAAAC TACTAAGGGGT CTAAGGGATC CTTAAAATGT TCCATATGG ATGCAGTATGA ATGCAGTATGA ATGTAAAATGT TCCATATGG TATGCAGAGA AGTTCAAATGT TCAGAGGACACA CCTCTATGAA ATGAGCACA TTCAGAGACAC TTCAGGAGACAC TTCAGGAGACAC TTCTGATAAC ATGAGACACAT TTAAGAGCAC AAAGAGCTAC AAACAGCTTCTC AGAGCCTTCTCACAAACACCACAC AAACAGCTTCTC	1111 1211 1411 1511 1611 1771 1811 2011 2211 2211 2311 2411 2511 2611 2911 2911 3011 3111 3311 3311
CHAMATOLI AGGACTANAA G GOTTOTNAGA G COTTGAGGA A COCTTGAGGA A COCTTGAGGA A ATGAGATAA A ATGAGATAA A ATGAGATAA A ATGAGATAA A ATGAGATAA A ATGAGACAA CHATTAAG G ATTATAGGA ACAGAAGTTA ACAGAAAGTTA ACAGAAAATA ACAGAAAGTTA ACAGAAAGTTA ACAGAAAATA ACAGAAAAAAAAAA	GLAATAACGA GCAAGGCCAT GCCTGAAGGG ACCTGAAGGG TTTGCTTAGA TTTGCTTAGA TTTGATAAGTG ACCAGACTTTA GGCATTTTAT GGCATTTTAT GCATTAAGGAG CTAAGAGAG TCATTTATA GCAGAAGCA TAGGTGAGTA TAGCCCGAGA ACCCCGAGA ACCCCGAGA ACCCCGAGA TGCCCCGAGA TGCCCCGAGA TGCCCCGAGA TGCCCCGAGA TGCCCCGAGA TGCCCCGAGA TGCCCCGAGA TGCCCCGAGA TGCCCCGAGA TGCCCCGAGA TTTAAAGCT TAAAACTCA TAAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAACTCA TAAAAACTCA TAAAACTCA TAAAACTCA TAAAAACTCA TAAAAACTCA TAAAAACTCA TAAAAACTCA TAAAAACTCA TAAAAACTCA TA	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATGCATT TATGCATT TATGCATT TATGCATT TATGCATA TAGCATCAT TAGCATCAT TAGCATCAT TACACTGAT TACACTGAT TACACTGAT TACACTGAT TACATTGAA TAGGACTTA AAATAAGAT TAGGAGCAT TAGGAGGACC TAGGAGGACC CCTGGAGCCT AAAGAGAATT AAATAATTCT TAGAATTATTCT AAATTATTCT AAATTATTCT AAAGGAATT ACACCTGTGA CTCTAGAGGA TTTGAACTT AAAGGAATT ACACCTGTGA TCTAGAGGATT AAAGGAATT ACACCTGTGA TCTAGAGGA TCTAGAGGA TCTAGAGGA TCTAGAGGA TCTAGAGGA TCTAGAGGA TGAGTTCCAC	TAGAGAGGTT TACATAGGTT TCATAAGGTT TAGAGATAGA TAGAGATAGA ATGATCCTT TAAAGCAGCA AGATGTCCTT TAAAGCAGCA ACAATTGGG GCTGATTTA ATTACTGGG GTCTGACTGC ACCATTGGG GTCTGACTGC TAGAGAAATTA ATTACTTGGG CTTGAGTAAATAT GCGAGAAATTA GAGAGATTTA GAGAAATTTA GAGAAATTA	CAGAAAGCTG GACATCTGGG TGGGGAAAGG TGGGGAAAGG AGATAATACT CAGTATTAAG CATTCAGACA TGAGGCCAGG AGAGAATAA CAAAGACTAA ATCAGATAAT TGCATTCAGACA ATCAGATAAT ATCAGATAAT TATGAGAGAT TATTGTGGGA CAGAGTCTGA TATTGTTGGAC TTTTTTTTGTGAC TGGACATTAAA CCTGATTGTG AAACAGAAGA CTTATTTATAA CCTGATTGTG AAACAGAAGA TGATTCAAAC TGATTCAAAC TGATTCAAAC TGATTCAAAC TGATTCAAAC TGATTCAAAC	ACTGECCTTE AGGACALAGC ATAGATAAGA AAGGAGAAAG TAATATTAGC TTACTGAGTT TITTCATCTG GCTTTGCAGG TTTTTCATCTG GCTTTGCAAG ACGCATCAT ACTGCAACA AGGCTAGAAC ATCCTGAATG TATCTGAATA ATCCTGAATG TATCAGCACA ATCCTGAATG TATCAGCACA AGGCTAGACC AATAGCTCA AGGCCATCG AGGGCTATGG AGGATTAGTC ACGATAGTC ACGATAGT	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT ACAGTTTGG ACACTACGG CAATATATGT ACCCTAAGGAG ACCCTAAGGAG ACCTTAAGA ACGTGAGGT ACTTAGGAG TTTAGTATG GATGAAGAGT CGGAATGAG TTTAGTATG CGGACTAGA TGGTTCGGGAT CCGGACTAGA TGGTTCACTC AGACGATACAC AGACGATACAC CCGAATTAAC CCGAACTAGA TTTCATTATA AACATGAAG CCTAATTAC CCGATCTGAT CCGATCTGAT ATCATCAGA	TOTCANCTIT TCCAGACAT ANAGAMATI TITCTTCTAT ACTIGGAGCIA ANATITTCATCAT TCTGCAGCCA ANATITTGAAA TGCCAGGTAC GATAMATGT GATAMATGT GATAMATGT GATAMATGT TACCTTGCA GATAMATGT TACCTTGCA GATAMATGT TTACCTTGCA GATAMAGAT TTGGAGAGGT TTGGAGAGGT TTGGAGAGGT TTGCCCAGC GAGTAMAGAT TCTGCAGACAA AAAATGTGAT AAAATGTGAT ATAMAGTT TTAAAAGTCT TAAAAGTCT TAAAAGTCT TAAAAGTCT TAAAAGTCT TAAAAGTCT TAAAAGTCT TAAAAGTCT TAAAAAGTCT TAAAAAGTCT TAAAAAGTCT TAAAAAGTCT TAAAAATTGAAA	ATCTCAGCTG TGTTGGAAA ATAAGTGGTG TGGGGCTCTG TTGAGGCAG TTGGGTTCTA TTAAACAGT TTGTTCAGGG GCACTAGGTA CGTTCTCT TTTTGTATGG AAAATGCAAT TGAATGCAT TGATGTAA AGCATGGTA AACCCTTAG AAACTGCAT ACCATGGAA AACCCTTAG CCAAAGTTAC CCAAAGTTAC CCAAAGTTAC CTAAGGCCTT TTGAAAT ACCATGGAA AACCAGGATA ATCAAGTGC TTAGAAATTA TAAGGAAGA AAAAAGAAGT TTAAAGTAC TTAAGGATC TTAAAAGAATTA TAAGGAAGGA AAAAAGAAGT TTAAAGGTTC TTATTTTCAC TTTTTTTCAC TTTTTTCAC TTTTTTTCAC TTTTTTCAC TTTTTTTCAC TTTTTTCAC TTTTTTTCAC TTTTTTTT	TGAATCAGA CTCACAAGC AGATGATGA GAAATAATGA AAAGGAGTT TACTGAAAAGC TACAAGGGT TTCATATGA TCCATATGA TCCATATGA TCCATATGA TCCATATGA TCTATAAATGT TTCATATGA TATGAGGTAT TATGAGCACA AGTTCATATGA TCAGCACAGA TCAGCACAGA TCAGCACAGA TCAGCACAGA TTTAGAGCAC TCAGGAGGGT TTAGGACACA TTTAGAGCAC AACCACCACA TCAGCACACA TCAGCACACACA TCAGCACACACA TCAGCACACACACACACACACACACACACACACACACACA	1111 1211 1411 1611 1711 1911 2011 2011 2111 2211 2211 2311 2411 2511 2611 2711 3011 3011 3111 3211 3211 3311 3411 3611 3711
CHAMATOL  GOTTOTAMA G GOTTOTAMA G GOTTOTAMA A CCCTTCAGGA A CCCTTCAGGA A CCCTTCAGGA A ATGGGATAA A ATGGGATAA A ATAGGARANG ATTAGACAN G ATTATAGA A CCAAGTATA A ACTAGCACAG A ATTATAGGA A ATTATAGGA A ATTATATOC T GTCAACACA T TATATTOC T TATATTOCAT TATATCAGA	GAMATANCGA GCAAGGCCAT GCAAGGCCAG GCCTGAAGGG AGGGGCCAGG TTTGCTGAGA TTGTAAAGTG ACCAGACTTCA GCAATTTAT GCAATTTATA GCAATTTATA GCAAGACAC CATTTATAA GCAGACAC TAAGGAGA TTGTAAAGGAG TTAAGAGAC TTAAGAGAC TGATTTATAA GCAGAAGCAT TAGGTGAGTA CCTGGCAGTA AAAGCTGAG AAAAGCTGAG AAAAGCTGAG TGACCAGAAAC TTAAAGAGCA TTAAAGACAC TTAAAGACAC TTAAAGACAC TTAAAGACAC TTAAAGACAC TTAAAGACCAC TTAAAGACCAC TTAAAGACCAC TTAAAGACCAC GATCACACAC AAGCTCACAG AAGCTCACAG CCTCTCTTTATAA	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATTGCATT TATGCATC TATTGCATC TATTGCATC TATTGCATC TATTGCATC TATTGCATC TATTGCATC TACACTGATC GCAAAGAAGG TAAACTAGGC TACACTGATT GACTATTGCA TAGACTTA TGGCTCTTTA AAATAAGATC TTGGAACTC GTGGAGCCC GTGGAGCCC GCGGAACCAGA AAATTATTCT AAAGAGAATT ACACCTGTGA CTCTGGAGC TCAGAGGAC TCAGAGC TCAGAGGAC TCAGAGC TCAGAGC TCAGAGC TCAGAGC TCAGAGC TCAGAC TCA	TAGAGAGGTT TACAGAGGTT TCATAGGTT TACAGATOC TAGGATACCC TAAGGATAGA TTCATGCTGT TAAAGCAGCA ATCACTCGTG ACAGCTGGTC ACAGCTGGTC ACAGCTGGTC ACAGCTGGTC ACAGCTGGTC ACAGTGCTA ATTATCTGGG GCTCAGTATA TTAGCTCATT TAGGAGAAACCC ATAGAGTAGT TTAGGTCATT GGGAGAACCC ATAGAGTAT GGGAGAACCC ATAGAGTAT GAGAGATATT GAGAGATATT GAGAGATATT GAGAGATATT GAGAGATATT CACCACAGAT	CAGAAAGCTG AAACTAGGAA GACATCTGGG TGGGGAAAGG CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAGACATA ATGAGATAA CAAAGACTA ATGAGATAA TATGAAGATAA TATGAAGATAA TATGAAGGC GGAGACATAA TATGAAGGC CAGAGTCTGA CTGGTACAT TATTATAAAA CCTGATTGTG AAACAGAAGA CTTATTATAA CTGATTGTG AAACAGAAGA CTTATTATAAC TGATTCAAAC TGAATTCAAAC TGAATTCAAAC TGAATTCAAAC TGAATTCAAAC TGAATTCAAAC TGAATTCAAAC	ACTGECCTTE AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTATTATTGC TTATTATTGC GCTTTGCAAG GCAGCACATCT GCATTGCAAG GCATCTTACAAG GCATTGCAAG AGGCTAGAAC CAATTGTAGAT ATCCTGAATG CTTTTGTAAG CTTTTGTAAG CTATTATTAACACACGT AGGGTCAGGA AGGCTAGGAT AGGGTCAGGG AGGGTCAGG AGGGTCAGG AGGTTTGTAAG CTATTGTAAG CTATTTCTGTAGGGCCACG AGGTCAGGG AGGTCAGGG AGGTCAGG AGGTTAGTAA ACATTGTTAA CTATTTCTGT GCAGATCAAA	TGGTTCCCAC TGGTGAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT ACAGTTTCG ACACTACGG CAATATATGT GGTCTTAAAA ACGTGAGGT ACCTAAGGAG ACCTTAAGGAG ACCTTAAGGAG CTGAACAGTA GAGTCTTGGG ATTTAGTATTG GATCAACGGA GTGTTAGGAG TTTTAGTATTG GATCAAGGAG AAAAAGGGAG AAAAAGGGAG AAAAAGGGAG AAAAAGGGAG AAAAAGGGAG CCTAATTTAC ACCTGAATTTAC CCTGATCTGAT	TOTCANCTIT TCCAGGANTG ANAGANANTT TCTCTATA ACTIGGETT TCTGCAGCCA TAAACTTCA TGCAGGTAC GATANANTGT GATAGGTAGG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTGGATAGG GCAGTACAAA GACTTTGGT TTACCTTGGA TTGGAAAAGT TTGGAAAAGT TTGGAAAAGT TTGGCAGAG TTGTCCCAGC CAGTAATCAAAG GAGTAGGAGAG TTGTCCCAGC CAGTAGAGTAG	ATCTCAGCTG TGTTGGAMA ATAMCTCGTG TGGGGCCTGG TTGGGCTTCTA TTAMACAGG CCACTAGGTA CGTTCTCT TTTTGTATCG AMATGCAGT AMATGCAGT TGGGCGGC AMAGGGGGC AMAGGGGGC AMAGGGGGC AMAGGGGGC TGGGTTAC AMACTGMAT AGCATTGC ATCTGAGT ATTCAAGGTC TTAGGAGTA ATCAAAGTTAC ATGGAATTA ATTAAGGTC TTAGGAGTA AMAAGAGAT ATAAGGTC TTAAGGTTC TTAAGGTTC TTAAGGTTC TTAAGGTTC TTAAGGTTC TTAAGGTTC TTTTTTAC ATCGTTCGAT	TGAATCAGA CTCACAAGC AGATGATGA AGATGATGA AAAGGAGTTT TCTGTAAAAGC CTTTAAATGT TTCCATATGG ATGGAGGTT TCTGACAAGC ATGGAGGTAT CGTACCAAGA AGGTCATATGA CCTTAAATGT TTCAATTGG TTCAATTGG TAGGAGGATAT TCAGGACAC TGTGATAACC AGGGCATAAA ATATATAAGG TGGATGGATAA ATATATAAGG TGGATGGA	1111 12111 1411 15111 1611 17711 1911 20111 22111 22111 22111 2311 2411 2711 2911 3011 3111 3211 3311 3411 3511 3611 3611 3711
CHAMATOL  GOTTOTAMA G GOTTOTAMA G GOTTOTAMA A CCCTTCAGGA A CCCTTCAGGA A CCCTTCAGGA A ATGGGATAA A ATGGGATAA A ATAGGARANG ATTAGACAN G ATTATAGA A CCAAGTATA A ACTAGCACAG A ATTATAGGA A ATTATAGGA A ATTATATOC T GTCAACACA T TATATTOC T TATATTOCAT TATATCAGA	GAMATANCGA GCAAGGCCAT GCAAGGCCAG GCCTGAAGGG AGGGGCCAGG TTTGCTGAGA TTGTAAAGTG ACCAGACTTCA GCAATTTAT GCAATTTATA GCAATTTATA GCAAGACAC CATTTATAA GCAGACAC TAAGGAGA TTGTAAAGGAG TTAAGAGAC TTAAGAGAC TGATTTATAA GCAGAAGCAT TAGGTGAGTA CCTGGCAGTA AAAGCTGAG AAAAGCTGAG AAAAGCTGAG TGACCAGAAAC TTAAAGAGCA TTAAAGACAC TTAAAGACAC TTAAAGACAC TTAAAGACAC TTAAAGACAC TTAAAGACCAC TTAAAGACCAC TTAAAGACCAC TTAAAGACCAC GATCACACAC AAGCTCACAG AAGCTCACAG CCTCTCTTTATAA	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATTGCATT TATGCATC TATTGCATC TATTGCATC TATTGCATC TATTGCATC TATTGCATC TATTGCATC TACACTGATC GCAAAGAAGG TAAACTAGGC TACACTGATT GACTATTGCA TAGACTTA TGGCTCTTTA AAATAAGATC TTGGAACTC GTGGAGCCC GTGGAGCCC GCGGAACCAGA AAATTATTCT AAAGAGAATT ACACCTGTGA CTCTGGAGC TCAGAGGAC TCAGAGC TCAGAGGAC TCAGAGC TCAGAGC TCAGAGC TCAGAGC TCAGAGC TCAGAC TCA	TAGAGAGGTT TACAGAGGTT TCATAGGTT TACAGATOC TAGGATACCC TAAGGATAGA TTCATGCTGT TAAAGCAGCA ATCACTCGTG ACAGCTGGTC ACAGCTGGTC ACAGCTGGTC ACAGCTGGTC ACAGCTGGTC ACAGTGCTA ATTATCTGGG GCTCAGTATA TTAGCTCATT TAGGAGAAACCC ATAGAGTAGT TTAGGTCATT GGGAGAACCC ATAGAGTAT GGGAGAACCC ATAGAGTAT GAGAGATATT GAGAGATATT GAGAGATATT GAGAGATATT GAGAGATATT CACCACAGAT	CAGAAAGCTG AAACTAGGAA GACATCTGGG TGGGGAAAGG CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAGACATA ATGAGATAA CAAAGACTA ATGAGATAA TATGAAGATAA TATGAAGATAA TATGAAGGC GGAGACATAA TATGAAGGC CAGAGTCTGA CTGGTACAT TATTATAAAA CCTGATTGTG AAACAGAAGA CTTATTATAA CTGATTGTG AAACAGAAGA CTTATTATAAC TGATTCAAAC TGAATTCAAAC TGAATTCAAAC TGAATTCAAAC TGAATTCAAAC TGAATTCAAAC TGAATTCAAAC	ACTGECCTTE AGGACAAAGC ATAGATAAGA AAGGAGAAAG GCAGCACTCT TTATTATTGC TTATTATTGC TTATTATTGC GCTTTGCAAG GCAGCACATCT GCATTGCAAG GCATCTTACAAG GCATTGCAAG AGGCTAGAAC CAATTGTAGAT ATCCTGAATG CTTTTGTAAG CTTTTGTAAG CTATTATTAACACACGT AGGGTCAGGA AGGCTAGGAT AGGGTCAGGG AGGGTCAGG AGGGTCAGG AGGTTTGTAAG CTATTGTAAG CTATTTCTGTAGGGCCACG AGGTCAGGG AGGTCAGGG AGGTCAGG AGGTTAGTAA ACATTGTTAA CTATTTCTGT GCAGATCAAA	TGGTTCCCAC TGGTGAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT ACAGTTTCG ACACTACGG CAATATATGT GGTCTTAAAA ACGTGAGGT ACCTAAGGAG ACCTTAAGGAG ACCTTAAGGAG CTGAACAGTA GAGTCTTGGG ATTTAGTATTG GATCAACGGA GTGTTAGGAG TTTTAGTATTG GATCAAGGAG AAAAAGGGAG AAAAAGGGAG AAAAAGGGAG AAAAAGGGAG AAAAAGGGAG CCTAATTTAC ACCTGAATTTAC CCTGATCTGAT	TOTCANCTIT TCCAGGANTG ANAGANANTT TCTCTATA ACTIGGETT TCTGCAGCCA TAAACTTCA TGCAGGTAC GATANANTGT GATAGGTAGG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTGGATAGG GCAGTACAAA GACTTTGGT TTACCTTGGA TTGGAAAAGT TTGGAAAAGT TTGGAAAAGT TTGGCAGAG TTGTCCCAGC CAGTAATCAAAG GAGTAGGAGAG TTGTCCCAGC CAGTAGAGTAG	ATCTCAGCTG TGTTGGAMA ATAMCTCGTG TGGGGCCTGG TTGGGCTTCTA TTAMACAGG CCACTAGGTA CGTTCTCT TTTTGTATCG AMATGCAGT AMATGCAGT TGGGCGGC AMAGGGGGC AMAGGGGGC AMAGGGGGC AMAGGGGGC TGGGTTAC AMACTGMAT AGCATTGC ATCTGAGT ATTCAAGGTC TTAGGAGTA ATCAAAGTTAC ATGGAATTA ATTAAGGTC TTAGGAGTA AMAAGAGAT ATAAGGTC TTAAGGTTC TTAAGGTTC TTAAGGTTC TTAAGGTTC TTAAGGTTC TTAAGGTTC TTTTTTAC ATCGTTCGAT	TGAATCAGA CTCACAAGC AGATGATGA AGATGATGA AAAGGAGTTT TCTGTAAAAGC CTTTAAATGT TTCCATATGG ATGGAGGTT TCTGACAAGC ATGGAGGTAT CGTACCAAGA AGGTCATATGA CCTTAAATGT TTCAATTGG TTCAATTGG TAGGAGGATAT TCAGGACAC TGTGATAACC AGGGCATAAA ATATATAAGG TGGATGGATAA ATATATAAGG TGGATGGA	1111 12111 1411 1511 1611 1711 1811 2011 2211 2211 2211 2311 2311 2311 3011 30
CHAMATOLI  GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G COTTGAGGA  COCTTGAGGA  ACCOTTGAGGA  ATAGGATAA  ATAGGATAA  ATAGGATAA  CHAGTATAA  CHAGTATAA  CHAGCACAA  ATATAAGAA  CHAGTATAA  CHAGCACAA  ATATAAGAA  COCAATAATA  CAGTAATAA  ACAGAAGTTA  ACAGAAGTTA  ACAGAAGTTA  CAGTAATAA  ACAGAAGTTA  CAGTACACA  GATTACACA  AGATGATTA  AGATGATTA  AGATGATTAA  AGATGATTA	GLAATAACGA GCAAGGCCAT GCAAGGCCATGA GCGGCCAGGGCCAGG FTGTAAAGGG FTGTAAAGGG GACACATTA GGCATTTATA GGCATTTATAA GTAAAGGAG CTAAGGAGAC CTAAGGAGAC CTAAGGAGAC CTAAGGAGAC CTAAGGAGAC TAAGGAGAC TAAGGAGAC CTAAGGAGAC TAAGGAGAC TAAGGTGAGA AAGGTGAGA AAGGTGAGA AAGGTGAC GATCACAGG CATAAAGAC TTAAAGAC TTAAAAGAC TTAAAAGAC TTAAAGAC TTAAAAGAC TTAAAAGAC TTAAAGAC TTAAAAGAC TTAAAAAAAAAA	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATGCATT TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC TACACTGAT GACTATTGCA TACACTGAT TACATATACA TTTGGAACTTC GTGGTGTAT AAATAAGACT TTTGGAACTC GTGGTGTAT AAGTAACAC GTGGAGCCT TACATATACA TTTGGAACTC GTGGTGTAT AAATAAGAC TTTGGAACTC TTTGGAGCC TACACTGGAGCC TACACTGGAGCC TACACTGGAGCC TACACTGGAGCC TACACTGGAGCC TACACTGTGA CCCTGTGA TCACTGTGA TTCACTGTGA TTCACTGGAT TTCACTGCAC TTCACTGGAT TTCACTGCAC TTCACTGCA	TAGAGGGTT TACAGAGGTT TCATAAGTGT TACAGATACCC TAAGGATACAC TAGAGTATACACT TAAGATACAC AGATGTCCTT TAAAGCAGCA ATCACTCGTG AGACTCGTC AGACTGGTC ACCACTGATTA ATTACTCGC TACAGCACTGCT ACCACTGCTATAAATAC ACCACAGTATAAATAC ACCACAGCACT TTAGACTCATTAAATAT GGCAGAAATTTA ATTACACTCATTAAATAT GAGAGAATTTA GAGAGAATTTA GAGAGAATTTA GACCACAGAT ACCACAGAT TTAGGTTAAAAAA TTAGGTTAAA	CAGAAAGCTA GACATCTGGG TGGGGAAAGG TGGGGAAAGG TGGGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG GAGACTATAA TATCAGATATA TATCAGATATA TATCAGATATA TATCAGATATA TATTCAGACA TATCAGATAC TGAGCCCAG GGACACTAA TATCAGATAC TGTCCCTAC TGTCCCTAC TGTCCCTAC TGTCCCTAC TGTACTTAAAC TGATTCAAAC TCAATTAAC TCAATTAAT TCCAATTAAC TCCAATTAAC TCCAATTAAC TCCAATTAAC TCCAATTAAT TCCAATTAAC TCCAATTAAC TCCAATTAAC TCCAATTAAT TCCAATTAAC TC	ACTGGCCTTG AGGACAAAGC ATAGATAAGA AGGAGAAAGG ATAGATATGATTTTTGAGTT TTACTGAGTTTTTTTCATTG GCTTTGCAAG GCAGCACTCTT GCATTGCAAGA AGGCTAGAAC GAAATGGCTA AGGCTAGAAC TACAGCACGT AACTGGAACA AGGCTAGAAC AACTGGAACA AACTGGAACA AACTGGAACA AACTGGAACA AACTGGAACA AGGCCATGGT AGGGCCATGGT AGGGCCATGGT AGAGTTACCT ACAATTAGTCA AACTAGGACA CAACTATTTA AAGCCATTTTT AAGCCATTTTT AAGCCATTTTT AAGCCATTTTT AAGCCATTTTTA AAGCCATTTTA AAGCCATTTATA AAGCCATTTTA AAGCCATTTTA AAGCCATTTTA AAGCCATTTATA AAGCCATTTTA AAGCCATTTATA AAGCCATTTATA AAGCCATTTATA AAGCCATTTATA AAGCCATTTATA AAGCCATTATATA AAGCCATTTATA AAGCCATTTATA AAGCCATTTATA AAGCCATTATA	TGGTTCCCAC TGGTGAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGGTAT AACAGTTTGG ACAGTTTGG ACAGTTAGA ACAGTTAGA ACAGTTAGA ACGTGAGGT ACCTAAGGAG ACCTAAGGAG CGGTATGAACAGTA GAGTCTTGGG TTTTAGTATG GATGAAGGAG AGTTAGGAG GTGTAGGAGTAGA CGGAACAGG AGTTAGGAG AAAAGGGAG TTTCATTAGTAG AAAAGGGAG AAAAGGGAG ATTTATTAAACAGAAAG CCTAATTAAC CTGATCTGAT	TOTCAGTTT TCCAGGATAG ANAGANANTT TTCTTCTAT ACTIGGETAT TCTGCAGCA ANTITGANA TGCCAGGTAC GATANATGT GATGAGTAGG ATTGGACAA ATTGGACAA ATTGGACAAT TGCAGGAGAT TTACCTIGCA GATANATGT GATAATGGACAAT TTACGAAGAGT TTACCTIGCA GATAATGGATAGGAGAT TTGGANAGAT TGGANAGAT TGGANAGT TGGANAGT TGGATAGGAT AGATCAACAA AGATCAACAA AGATCAACAA AGATCAACAA AGATCAACAA AGATCAACAA ATTAGACAA ATTAGACAA ATTAGACAA ATTAGACAA ACATTGTATT ACTTGGACAA TCTAAAGTCT AAAAAAATGGA ATTAGACAA TCTAAAGTCT AAAAAAATGGA ACATTGTATT ACTTGGACAA TTTAGACCAT TTTAG	ATCTCAGCTG TGTTGGAAA ATAAGTGGTG TGGGGCTCTG TTGAGGCAG TTGGGTTCTA TTAAACAGT TTGTTCAGG GCACTAGGTA ATAATGCAT TTTTGTATGG AAAATGCAT TGTTTGTATGG AAAATGCAT TGGAGGGGG AAAAGAGGGG TTGATGTATA AGCATTAGAAT AGCATGGATA AGCATGGATA ATCAAGTGC TTAGAAATTAAG TTAAGAATTA CTAAGAGCTG TTAAGAATTA AGCATGGATA ATCAAGTGC TTAGAAATTA AGCATGGATG ATAGAAGTTA CTAAGAGTT CTTAGGAGT ATAGAATTAAG AAAAGAGGT CTTTTTCAC ATCGTTCGAT ATCGTTCGAT ATCGACGTCT TTAAGACTT	TGAATCAGA AGATGATTA AAAGGATTT TCTGTAAAAG TACATGAGGGT TACATGAGGATT TCTGTAAAAG ATAGAGGATT TCTGTAAAAG ATGAGGATT TCTGAGAGGT TTCAAATGA AGTTCAAATG TATGAGGATA TATGAGGATA TATGAGCACA AGTTCAAATG TATGACACAGA AGTTCAAATG TCAGGAGACAGT TCAGGAGACAGT TCAGGAGACAGT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACT TCAGGAGACACT TCAGGAGATAAAAG ACAGCTACT TCAACCACCAC TCTACTGACACCAC TCTACTGACACCACC TCTACTGACACCACC TCAACCACCACC TCAACCACC TCAACCACCACC TCAACCACC TCAACCACCACC TCAACCACC TCAACCACCACC TCAACCACCACC TCAACCACCACC TCAACCACCACC TCAACCACCACCACC TCAACCACCACCACC TCAACCACCACCACC TCAACCACCACCACC TCAACCACCACCACC TCAACCACCACCACC TCAACCACCACCACCACCACCACCACC TCAACCACCACCACC TCAACCACCACCACCACCACCACCACCACCACCACCACCA	1111 1211 1411 1511 1611 1771 1911 2011 2011 2211 2211 2211 2211 221
CHAMATOL AGGACTANA AGGACTANA AGCOTTCTAGGA ACCOTTCTAGGA ACCOTTCTAGGA ACCOTTCTAGGA ACCOTTCTAGGA ACCOTTCTAGGA ACCOTTCTAGA ACCOTTCAGA ACCOTTCAGA ACCOTTCAGA ACCOTTCAGA ACCOTTCTAGA ACCOTTCAGA ACCOTTCAGA A	GLAATAACGA GCAAGGCCAT GCCTGAAGGG ACCTGAAGGG ACCTGAAGGG FTTGATAAGTG ACCAGATTTAA GCAATTTATA GCAATTTATA GCAGTCCTGA ACATAAGGAG ACATAAGGAG ACATAAGAGAC TATATATAA ACAGAAGAT TATATTATAA ACAGAAGCAT TAGGTGAGAA CCTGGCAGA ACATAAACTGA GAGCAAAGGG GATCAAGCATA TTAAAGAGCTTA TTAAAGAGCTTA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CATAAACTGA CTCTTTTTATC MAGTACAAAGG CTCTTTTTATC MAGTACAAAGG GCACCAATTATCA	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TATTGCATT TATGCATCA AATGCGGAAA CCCTTTGATC AATGGGGAAA GCAMAGAAGG TAACCACTGAT GACTATTGCA GTATGACTTA TGGACTTTA AATAAGAC GCGGGAACCAGA AATTATCCA GCGGAACCAGAC GCGGAACCAGA AATTATTCT AAGAGACT GCGGAACCAGA AAATTATCT AAGAGAATT ACACCTGTGA AATTATTCT AAAGAGAATT ACACCTGTGA CTCAGAGGAC TTCATTATT CGAGCTGTTCATTATT CGAGCTGTTTATT TCATTATTTCT TTCATTATTTT TGAGCTGTTTTCATTATT TGAGCTGTTTTCATTATTTTCATTATTTCT TAGAGGATGTGTTTCATTATTTCTTCATTATTTTCATTATTTTCATTATTTTCATTATT	TAGAGGGTT TACAGAGGTT TCATTAGGTT TACAGATCCC TAGGATCCCC TAGGATCCCC TAGGATCCCC TAGGATCCCC TAGGATCCTC TAAAGCAGCA CAATTGATAG ATCATCCGTG GGCTCATTTA ATTATCTCCTG GGCTCATCTA ATTATCTCCTT TAGGTATA ATTATCTCTT TCACCCACAGAT TTAGGTATA ATTATCTCTT TGCTCACTGGC ATAGATTTA ATTAGATATA TCATTAGATATA TAGGATCAT TAGGATCAT TAGGATCAT TAGGATCAT TAGGATCAT TAGGATATA ATTAGGATAT ACCCACAGAT TAGGTATA ACTCAGGAACA ACCCCTAAACAA TTAGGTATA ACTCAGGAACA TTAGGTATA ACTCAGGAACA ACCCCTAAACAA TTAGGTATAA ACTCAGGAACA ACTCAGAGAT ACTCAGGAACA ACTCAGGAACA ACTCAGAGAT ACTCAGGAACA ACTCAGAGAT ACTCAGGAACA ACTCAGAGAT ACTCAGGAACA ACTCAGAGAT ACTCAGAGAACA ACTCAGAGAT ACTCAGAGAACA ACTCAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAACA ACTCAGAGAACA ACTCAGAGAACA ACTCAGAACA ACTCAGAACA ACTCAGAACA ACTCAGAACA ACTCAGAACA ACTCAGAACA ACTCAGAACA ACTCAGAACA ACTCAGAACA	CAGAAAGCTG AAACTAGCAA GACATCTGGG TGGGGAAAG TGGGGTATAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG TGAGGATAAT CAAAGACTCA AAACACTCA GAAGAATAT GGATGCCCAG GAAGACATAA TATGGAAGAG CTGGCTACAT TATTATACA TGTATTTTTTTTTT	ACTGECCTTE AGTGGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAGC TTATTATTGC TTACTGAGTT TITTCATCTGAGTT TITTCATCTGAGTT TITTCATCTGAGTT TCTGCAACA AGGCTAGAAC TCTGCCAACA AGGCTAGAAC TACTGGTATAAC CAAATAGCTA ATGCCTTTGTAG TACTGATAGAC TACTGCAACA AGGCTCAGGG ATGCCTTGTT AGGGCCATGG AGGGTTAGAC AGGCTTAGTAG CAATATCCT ACAATAGCAC TACTGTAAA CCAATTTCAG GCAGCATGG GCAGCATGG AGGTTTCCT ACAATAGCAC TACTGTTAAA AAGTAGTCA AAGTATTTAAA AAGCACATTTT GGAGTCAGGG CAACTATTTC GGAGTCAGGG CAACTATTTC GGAGTCAGGG CAACTATTTC GGAGTCAGGG CAACTATTTC GGAGTCAGGG CAACCATTTT GGAGTCAGGC CAACCATTTT GGAGTCAGCC CAACCATTTT GGAGTCAGGC CAACCATTTT GGAGTCAGGC CAACCATTTT GGATCAGCC CAACCATTTT GGAGTCAGCC CAACCATTTT GGATCAGCC CAACCATTTT GGATCAGCC CAACCATTTT GGATCAGCC CAACCATTTT GGATCAGCC CAACCATTTT GCAACCATTTT GCAACCATTT GCAACCATT GCAACCATTT GCAACCATTT GCAACCATT GC	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTAG ACAACTACGG CAATAATGT ACTATTAGAA AACAGTAAGGA CGCTAATGAA ACGTAAGGA CGCTAATGAA AGGTTAGGA TGTTAGGAG TTTAGTATG TGATCACTC AGAACATTAGAT AGGTTAGGAG TTTAGTATG TGGTTCAGCA AGGTAACAT AACATTAACATT AACATTAAAAG CCTAATTAAA CCTAATTAAA CCTAATTAAA CCTAATTAAA CATAGCATGAA TGGATGAAA AACATGAAAG CATAGCATT TATATCAGGA ATGGATGCAA ATG	TOTCANCTIT TCCAGGARTE ANAGAMANTI TITCTTCTAN ACCTGGETT TCTGCAGCCA ANATTTGANA TGCCAGGTAC GATANANTGT GATGAGTAGG ATTTCAGGAG ATTTCAGGAG ATTTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG TACCTTGTGT TTACCTTGCT TTACCTTGCA GACTTTGCAAA GACTTGCAAA CACTTGGTAAG TTGCCCAGC GAGTAACAA AGATCAACAA TAAATGTGAAACAA TAAATGTGAAACAA TAAATGTGAAACAA TAAATGTGAAACAA TAAATGTGAAACAA TAAATGTGAAACAA TAAATGTGAAACAA TAAATGTGAAACAA TAAATGTGAA TCTAGAGAGT TCTAGGAGAG TCTAGGAACAT TAAAAAATGTGA TAAAAAAATGTGA TAAAAAAATGTGA TAAAAAAATGTGA TAAAAAAATGTGA TAAAAAAATGTGA TATTAGGAGCAT TTTAGGAGCAT TTTAGGAGCAT TTTAGGAGCAT	ATCTCAGCTC TCTTCGANAA ATAACTCCTG TCGAGCCATCC TTGTAGCCAG TTGATCAGGG TCGACTACGTA TTAAAACAGT TCTCTCAGGG CCACTAGGTAT AAAATGCAT CCACAGGCCCC AAAAGAGGGG CCAAAGTTAC AACTGCAAAT AGCACTTAGA AACCCCTTAG CCAAAGTTAC AATCTGAAAT AGCACTGAAA TTAAAGTAC ATTAAAGTAC CTTCTTCAAT AGCTTCCAT AGCTTCAAT AGCTATTAAAG CCACAGAGTCT CCAAGAGTCT CCAAGAGTCT CCAAGAGTCT	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA AAAGGAGTTT TCTGTAAAAC TACTAGGGGT TTAGGAGTAT TCCATATGG TTCCATATGG TTCCATATGG TTCAGAGTAT CGTACAAGA AGTTCAATGT TCCATATGG TATGGAGTAT CGTACAAGA AGTTCAAATGT TCAGGAGTAAA AGTATCAAATGT TCAGGAGCACA TTCAGGAGCACA TTCAGGAGCACA TTCAGAGACACA TTTAGAGCACA TCAACACAC TCCACACAC TCCACACAC TCCACACAC TCAACACACAC	1111 12111 1411 15111 16111 17111 19111 20111 22111 22111 22111 22111 23111 23111 30111 31111 35111 35111 35111 36111 3711 36111 3711 36111 3711 36111 3711 36111 3711 36111 3711 36111 3711 37
CHAMATOLI AGGACTANAA G GOTTOTAMAA G COTTOTAMAA G COTTOTAMAA COTTOTAMAA COCTTOTAMAA ACCOTTOTAMA ATMAGGATAA ATMAGGATAA ATMAGGATAAA COLAGATATAA COLAGAGAAA COLAGAGAAA COTTOTAMAA COLAGAGAAA COTTOTAMAAA COTTOTAMAAA COLAGAGAAA COLAGAGAAAA COLAGAGAAAA COLAGAGAAAA COLAGAGAAAA COLAGAGAAAA COLAGAGAAAA COLAGAGAAAA COLAGAGAAAAA COLAGAGAAAAA COLAGAGAAAAA COLAGAGAAAAA COLAGAGAAAAAA COLAGAGAAAAAA COLAGAGAAAAAA COLAGAGAAAAAAA COLAGAAAAAAAA COLAGAGAAAAAAAA COLAGAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	GLANTACCA GCAAGGCCAT GCATGAAGGC ACCTGAAGGG ACGGGCCAGG ACGGGCCAGG ACCAGACTTA GGGATTTAT GGCATTTATA GCAGACTCC ACATAAGGAG ACATAAGGAG ACATAAGGAG ACATAAGGAG ACATAAGGAG ACATAAGGAG ACATAAGAGAC CTAAGAGAC CTAAGAGAC CTAAGAGAC CTAAGAGAC CTAAGAGAC CTAAGAGAC CTAAGAGAC CTAGAGAC CTAAGAGAC CTAAGAGAC CTAGAGAC CTATTATAA CCAGAAGC CTATAAAC CCTGGC CATAAACC CACAATT CAGACCAATT CAGACCAATT CAGACCCAATT CAGACCCAATT CAGACCCAATT CAGACCCAATT CAGACCCAATT CAGACCCAATT CAGACCCCAATT CAGACCCCAATT CAGACCCAATT CAGACCCCAATT CAGACCCCCAATT CAGACCCCCAATT CAGACCCCCAATT CAGACCCCCAATT CAGACCCCCAATT CAGACCCCCAATT CAGACCCCCCCC CAGACCCCAATT CAGACCCCCAATT CAGACCCCCCCCCC	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TATGCATT TATGCATT TATGCATT TATGCATT TATGCATT TATGCATT TATGCATC TATGCATC TATGCATC TACACTGAT GCATATAGC TACACTGAT GACTATTGCA TACACTGAT TACACTGAT TACACTGAT TACACTGAT TACACTGAT TACACTGAT TACACTGAT TACACTGAT TACACTGAGCA TACACTGAGCA TACACTGAGCA TACACTGAGCA TACACTGAGCA TACACTGAGCA TACACTGAGCA TACACTGAGCA TACACTGAGCA TACACTGAGAT TACACTGAGT TACACTGAGAT TACAC	TAGAGGGTT TACAGAGGTT TCATAAGGTT CTTCAGATGC TAAGGATAGA TACAGTCCT TAAGGATAGA AGATGTCCTT TAAGGATAGA ATCATTCGTG AGATGTCCTT ATTATTCTGTG ACCATTGGG TCACTGGGT ACCATTGGG ACCATTGATA ATTATTCTGTG ACCATTGCTA ATTATTCTGTG ACCATTGCTA ATTATTCTGTG ACCATTGCTA TTAGGTATAA TTAGGTATAA TTAGGTATAT GAGAAATTTA GAGAAATTTA GAGCAGATACA TAGGTATAT TAGGTTATAT ACCACAGGT CCCTAAACAA TTAGGTTATAT ACCACAGGT CCCTAAACAA TTAGGTTATATA ACTCAGGAAC GAGGGAAC GAGGGAAC GAGGGAAC GAGGGAAC GAGGATACAA TTAGGTTATATA ACTCAGGGAAC GAGGGAAC GAGGGAAC GAGGGAAC GAGGGAAC GAGGGAAC TTAGGTTATATA ACTCAGGGAAC GAGGCATCAA	CAGAAAGTCA GACATCTGGG TGGGGAAAGG TGGGGAAAGA AGATAATACT CAGTATTAAG TGCCTTTAGG CATTCAGACA TGAGGCCAGG AGAGAATAA CAAAGACTCA ATCAGATAAT CGATGCCCAG GGAGACATAA TATTGAGAGG CAGAGTCTGA TTTTGTTGGA TTTTGTTGGA TTTTGTTGAC TGTGCCCTAC TGCACATTAA TGCACATAAA TGCACATAAA TGCACATAAA TATTTTACTGGTGG GACTGTAAAT TATTGTGGTGG	ACTGECCTTE AGGACALAGC ATAGATAAGA AAGGAGAAAGC TTAATATTGC TTACTGAGTT TTACTGAGTT TTACTGAGTT TTACTCTGAGTT TTACTCTGAGTT TTACTCTGAGTT TTACTCTGAGTT TTACTCTGAGTT TTACTCTGAGT GCTTTGCAAG GAAGTCATTA ATCCTGAATAG ATCCTGAATAG TACAGCAGT ATGCAGAGAG AGGACTACGA CAATAGTCA AAGTAGGCC TACTGTTAAA CTACTTATA CAACTATTTA AAGCCATTTTA AAGCCATTTTA AAGCCATTTTA AAGCCATTTTA AAGCCATTTTA AGGCCATGCGC CATCCATCGGC CATCCATCGGC CATCCATCGGC	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT GCAATGTGAT ACAACTTACG ACACTACGG ACACTACGG ACACTACGG ACCTAAGGAT ACCGTAAGGAT ACCGTAAGGAT ACGTCAAGGAT AGACAGTAGA TTTAGTATG GATCAAGGAC TGGTTGGGAT CCGGAACAGA CCGAACAGA CCGAACAGA CCGAACAGA CCGAACAGA CCGAACAGA CCGAACAGA CCGAACAGA CCGAACAGA TTTCATTCAT AACATTCAAG CCTAATTACA CCTAACAGTT TATAACGAG ATTGATCAG CATACCATT TATAACGAG ATGACACTT CATTACCAG CATTACCAG CATTACCAT CCTATTCGTC CATTACCAT CCTATTCGTC CATTACCAT CCTATTCGTC CATTACCAT CCTATTCGTC CATTACCAT CCTATTCGTC CATTACCAT CCTATTCGTC CCTATTCGTC CCTATTCGTC CCTATTCGTC CCTATTCGTC CCTATTCCTC CCTATTCGTC CCTATTCGTC CCTATTCGTC CCTATTCCTC CCTATTCCT CCTATTCCTC CCTATTCCT CCTA	TOTCANCTIT TCCAGACATE ANAGAMANT TITTCTTCATAT ACCTGGCTT RCTGCAGCCA ANATTTGANA TGCCAGGTAC GATANANTGT GATGAGATAG ATTTCAGAGA ATTGACATA ATGATAGTAG ATTGACATAG ATGATAGTAG GCAGTACAAA ATGATAGTAG GCAGTACAAA ATGATAGTAG TTAGCAGAGT TTAGCAGAGT TTAGCAGAGT TTAGCAGAGT TTAGCAGAGT TGCCCAGC GAGTAATC AGATCAACAA TAAATGTAT GAGTAGGGAG ATTTGACAAA TCTAAAGGAG TCTAAAGGAG TCTAAAGGAG ACATTGACAAA TCTAAAGGAG ACATTGACCAA TAAAAAATGTAT TTAAAGGAGT ACATTGACCAAA TTTAAAGGAGT ACATTGACCAAA TTTAGAGGT ATTTTTTCC ACAAACAGAGT ATTTTTTTCC ACAAACAGAGT ACAAACACACACACACACACACACACACACACACACAC	ATCTCAGCTG TGTGGAAA ATAAGTGGTG TGGGGCCTGG TTGGGCTTCTA TTAAACAGT TTGTCAGGG CCACTAGGTA CGTCTCTCT TTTTGTATGG AAAATGCAAT AGCACTTGTATAGC AAAAGAGGGG CAAAAGTGAA AAGCCCTTAG CCAAAGTTAC AAATGCAAT ACCACTGAAT ACCACTGAAT ACCACTGAAT ACAAGCCTGT CTTAGAAT ACAAGCCTGT CTTAGAAT ACAAGCCTGT CTTAGAAT AAAGAAGTAC ATTAAAGTAC ATGTTTCAGATTAC CTTAGAATTAAG AAAAGAAGTA AAAAGAAGT TTAAAAGATTAC TTAAAGTTCAT TTAAGGACCA TTAAAAGATTC TTTTTTTTTT	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTT TCTGTAAAAC TACTAGGGGT CTAAGGGATC CTTCAAATGT TCCATATGG TCCATATGG TCCATATGG ATGGAGGTAT TCCATATGG TATGGAGGTAT TCAGAGACACA AGTTCAAATGT TCAGAGACACA TCAGGACACA TCAGGACACA TCAGGACACA TCAGGACACA TCAGGACACA TTAGAGACAC TTGTGATACCAC AACACCAC AACACCAC AACACCAC AGAAAACTAGA AGAAACTAGA AGAAACTAGA AGAAACTAG TGATATTTTCC AGGATGGTTCT TCAGGAGACA AGAAACTAGA AGAAACTAGA AGAAACTAGA AGAAACTAGA TGATATTTTCC TCAGAAGAGACACAC AGGAAAACTAGA TGATATTTCC TCAGACAGAC AGGAAACTAGA AGAAACTAGA TGATATTTCC TCAGACAGAC AGGAAACTAGA AGAAACTAGA TGATATTTCC TCAGACAGAC AGGAAACTAGA AGAAACTAGA AGAAAACTAGA AGAAACTAGA AGAAAACTAGA AGAAACTAGA AGAAAACTAGA AGAAACTAGA AGAAACTAGA AGAAACTAGA AGAAACTAGA AGAAACTAGA	1111 1211 1411 1611 1711 1911 2011 2011 2011 2211 2211 2211 22
CHAMATOLI  AGGACATANA  GOTTOTHAGA  GOTTOTHAGA  CCCTTGAGGA  CCCTTGAGGA  ATGAGATAN  ATGAGATAN  ATGAGATAN  CTATTANA  GOTATANA  CTAGTATAN  CTAGTATAN  CTAGTATAN  CTAGTATAN  CTAGTATAN  CTAGTACAN  ATGAGACACA  ATTAGACACA  TATTAGACA  TATTAGACA  TATTAGACA  TATTAGACA  TATTAGACA  TATTAGACA  TATTAGACA  CTAGAAAAAG  CTAGAAAAAG  CCAGTAAATA  CCAGTAAATA  CCAGTAATA  CCAGTAATA  CCAGTAATA  CCAGTAATA  CCAGTAATA  CCAGTACACA  GATTACACA  GATTACACA  GATTACACA  GATTACACA  CAGTAGACA  GATTACACA  GATTACACA  GATTACACA  GATTACACA  CCAGTACACA  GCGAACAGAA  GCCAATGAACA  CCCATGACACA  CCCATGACA  CCCATGACACA  CCCATGACACA  CCCATGACACA  CCCATGACACA  CCCATGACACA  CCCATGACAC  CCCATGACAC  CCCATGACAC  CCCATGACAC  CCCATGACACA  CCCATGACAC  CCCATGACAC  CCCATGACAC  CCCATGACA  CCCATGAC  CCCATCAC  CCCATGAC  CCCATGAC  CCCATGAC  CCCATGAC  CCCATCAC  CCCATCAC	GLAATAACGA GCAAGGCCAT GCAAGGCCAT ACCTGAAGGG ACGGGCCAGG FTTGAAAGGG ACCAGACTTAA GGCATTTAA GGCATTTAT GGCATTTAT GCATTTAAGGG ACAGACTTAT ACGAGACTC CATAGGGAC CTAAGGGAC CTAAGGGAC CTAAGGGAC CTAAGGGAC CTAGGGAGAC CTAAGGGAGAC CTAAGGGAGAC TAAGGGAGAC TAAGGGAGAC CTACAGAGAC CTACAGACCAT ACGCTGAGA AAGGTGAGAC CATAAACCC CATAAACCC CATAAACGC CATAAACCC CATAAACC CATAAACCC CATAAACC CATAAACCC CATAAACC CATAAACCC CATAAACC CATAAACCC CATAAACC CATAAACCC CATAAACC CATAAACCC CATAAACC CATAAACCC C	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATGCATT TATGCATT TATGCATT TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC TACACTGAT GACTATTGCA TACACTGAT TACACTGAT TACACTGAT TACATTGAA TACACTGAT TACATTGAA TACACTGAT TACATTATACA TTTGGAACTC GTGGAGCCT TACATGACA TACACTGTAT AAATAAGATC TACACTGTAT TACATTATACA TTTGAACTC TTTGAACTC TTTGAACTC TTTGAACTC TTTGAACTT AAAGGAATTATTC TACATTGAACT TACATTGAACT TACATTGAACT TACATTGAACT TACATTGAACT TACATTGAACT TTTCATTGAACT TTTCATTGATT TACACTGTGA TTTCATTGATT TACACTGTGA TTTCATTGATT TACACTGTGA TTTCATTGATT TACACTGTGA TTTCATTGATT TACACTGTGA TTTCATTGGAT TACACTGCTGCA AGGCTCGT AGGAGCCTGC AGGCCTGCCA	TAGAGGGTT TACAGAGGTT TCATAAGGTT CTTCAGATOC TAAGGATAGA TTCATGCTGC AGATGCCCT TAAGGATAGA ACACTCGTG ACACTCGTG ACACTCGTG ACACTCGTG ACCACTCGTATTA ATTACTCGG CTCGACTACTATA ATTACTCGG CTCGACTATTA ATTACTCGT CCCCAGTAAATA TTACGTCATTA	CAGAAAGCTA GACATCTAGGA GACATCTAGGA TGGGGCAAAGC TGAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG GAGACATAA ATCAGATAAT TGCATTTAGG CAGAGACTAA ATCAGATAAT TATAGACATAA TATAGACATAA AAACAGTAGT TTTTTTGTGGA CAGAGTAGTAGT TATTATAAA CCTGATTGTG AAACAGTAGT TGATCAAAC TCAATTAAC TCACTAAAT TATTATAAC TCACTAAAT TACTGGTGG GGCTTAGAT TACTGCTGGG GGCTTAGAT TACTGCTGGG GGCTTAGAT TACTGCTGGG GGCTTAGAT TACTGCTGGG GGCTTAGAT TACTGCTGGG GGCTTAGAT TACTGCTGTGG GGCTTAGAT TACTGCTGTGG CACCTTATAT	ACTGGCCCTG AGGACAAAGC ATAGATAAGA AAGGAGAAAGC TTATTATATTGC TTATTATTGC TTATTATTGC TTATTATTGC GCTTTGCAAG GCAGCACTCT TTTTCCAGGTT TGAGTGCAACA AGGCTAGACA AGGCTAGACA ATCCTGAATG ATCCTGAATG ATCCTGAATG ATCCTGAATG ATCCTGAATG AGGCTATGCA AGGCTATGCT AGGGCCATGT AGGGCCATGT AGGATCATTATTA AGGTTAGAC CAATTATTTA AGGTTAGAC CAATTATTTA AGGCATTATTA AGGCATTATTA AGGCATTATTA AGGCATTATTA AGGCATTATTA AGGCATTATTA AGGCATTATTA AGGCATCTG TGCAGGGGC CATCAGGGGC CATCAGGGGC CCACCGGCACTTT TGCACGGACCACTTG TGCAGGACCACTTT TGCACGGACCACTTG TGCACGCACCTTC TTATTATTATATATATATATATATATATATAT	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT ACAGTTTCGCTAT ACAGTTTCG ACACTACGG CAATATAGGT GGTCTTAAAA ACGTGAGGT ACCTTAGGAG CCGTAATGAT GATCAGTAGA GTGTAGAACGTA GATCAGTAGA GTGTAGAACGTA GATCAGGAG GTGTGGGCAT CCGGACTAGA TGGTTCACTC AGACGATAGA TTTATTATA AACATGAAG ATTAATTATA AACATGAAAG CCTAATTTATA AACATGAAAG CCTAATTATA AACATGAAAG CCTAATTTATA AACATGAAAG CCTAATTTATA AACATGAAAG CCTAATTTAC CCTACTGAT TATATCGAGG CATAGCATT TATATCGAGG CATAGCAT TATATCAGAG CATAGCAT TATATCGAGG CATAGCAT TATATCAGAG CATAGCAT TATATCAGAG CATAGCAT TATATCAGAG CATAGCAT TATATCAGAG CATAGCAT TATATC	TOTCHACTIT TCCAGGATA AAAGAAAATT ACTGGCTT TCTGCAGCCA AATTTGAAA TGCCAGGTAC GATAAATGT GATGAGTAG ATTGGACAA ATTGGACAA ATTGGACAA TGCAGGTAC GATGAGTAGG ATTGGACAAT TACCTIGCA GATAATGTGG GAGTACAAA GACTTTGGAAGAT TTGGAAGAT TTGGAAGAT TTGGAAGAT CTGGAAGAT AGTTGATATCA AGATCAACAA TCTAAAGTCT AAAAAGTGAT AAATGTGAT TACTTGAAA TCTAAAGTCT AAAAAAATGT AATTGAATAT TTAAGGAGAT TTAAGGAGAT TTAAGGAGAT TTAAGGAGAT TTAAGGAGAT TTAAGGAGAT TTAAGGAGAT TTAAGGAGAT TTTAAGGAT TTTAGGAT TTAGGAT TTTAGGAT TTTAGAT TTTAGGAT TTTAGAT TTTAGGAT TTTAGAT TTTA	ATCTCAGCTG TGTTGGAAA ATAAGTGGTG TGGGGCTCTG TTGGGCTTCTA TTAAACAGT TTGTTCAGGG GCACTAGGTA ATAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT AGCATGGAT AGCATGGAT AGCATGGAT ATCAAGTGAC TTAGAAATTAATTAAGGAAGTA TTAAGGAAGTA TTAAGGAAGT TTAAGGAAGT ATCAAGTTC TTTTTTTTCAC GTTTTTTTTCAC GTTTTTTTTCAC GTTTTTTTTAAG CCACAGATCT TCCAGGGCCA TTCACAGGCCCA TTCACAGGCCCA TTCACAGGCCCA ATCGTTCAC	TGAATCAGA CTCACAAAGC AGATGATGA AGATGATGA AAAGGAGTT TTCTAAAAGC CTAAAGGGGT CTAAAGGGTT TTCCATATGA CGAACCAAGA AGTTCAAAAGC ATGGAGGTAT TTCAATAGGGT TTCAATAGG TATGAGGATA AGTTCAAATG TATGACCAC AGGGACAGT TCAGGAGACAGT TCAGGAGACAGT TCAGGAGACAGT TCAGGAGACAGT TCAGGAGACAGT TCAGGAGACACAC AACACCACCAC AACACCACCAC AACACCACC	1111 1211 1411 1511 1611 1711 1911 2011 2211 2211 2211 2311 2411 2511 2611 2911 3011 3311 3311 3311 3411 3511 3711 4011 4011 4011
COLAMATOL A GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G COTTGAGGA A COCTTGAGGA A COCTTGAGGA A ATAGGAGAA A ATAGGAGAA A ATAGGAGAA A ATAGGAGAA A ATATGACA G CGAGTATA A CGAGTATA A CGAGTATA A CTAGGACAA G ATATAGGA A ATATAAGGA A ATATAAGGA A ATATAAGGA A TATTATOT T TATATTO T T TATATTO T T TATATTO T T T T T T T T T T T T T T T T T T	GAAATAACGA GCAAGGCCAT GCAAGGCCAGGCCAGGCC	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATGCATT TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC GACTATTGCA GCAAAGAAGG GCAAAGAAGG GCAAAGAAGG TACACTTGAT TGCATCTTTA TGGCTCTTTA AAATAAGATT TTGGAACTTC GTGGTGTAT AAGTAACA TTTGGAACTC GTGGTGTAT AAGTAACA TTTGGAACTC GTGGTGTAT AAATAAGA TTTCATGAACT TTTGAACTC TTTGAACTC TTTGGAACTC TTTGAACTC TTTGAACTC TTTGAACTC TTTGAACTC TTTGAACTC TTTCATGATT TTCCATGCAT TTCCATGCAT TTCCATGCAT TTCCATGCATC TTCCATC TTCCATGCATC TTCCATGCATC TTCCATGCATC TTCCATGCATC TTCCATGCATC	TAGGAGGTT TACAGAGGTT TACAGAGGTT TACAGAGGTT TACAGATCCC TAAGGATAGA TACAGCTGCT TAAGCAGCA AGATGTCCTT TAAAGCAGCA ATCACTCGTG AGAGCTGGTC ACACTGGTG ACCATCTGGG TCTGACTGC ACCATCTGGT ACCATCTGTT AATTACTCGCT ACCAGTGCTATA ATTACTTCGT TACAGTGCTATA ATTGTTCTGT GCTCAGTAAAAT TAGGTGATTAGGAGAACCC ATAGAGTGAT TTAGGTCATT GAGAAATTTAG GAGAAATTTAG TAGGTAAACAA TTAGGTCATT TAGGTAATAT TAGGTATATA ACTGAGGAAC TTAGGTATATA ACTGAGGAAC TTAGGTATATA ACTGAGGAAC TTAGGTATATA ACTGAGGAAC TTAGGTATATA ACTGAGGAAC TTAGGTATATA TTAGGTTATA ACTGAGGAAC TTAGGTATATA TTAGGTTATA TTAGGTTATA TCCATATTAG TGGTGGAC TGGAGGAAC TGGAGGAACCC GTGAAACCC	CAGAAAGCTA GACATATTAGA GACATATTAGA GACATATTAAT TGCCTTTTAGG CATTCAGACA TGAGGCCAGG GAGACATATAAT CAATATAAT CAATATAAT CAATATAAT CAATATAAT CAAAGACTCA ATCAGATATAAT TGCATTGGAGA GAGACCTAA TATGGAAGGG CACAGTCTGA AACAGTAGT TTTTTTTGAG TGTCCCTAC TGGACATTATA TATTATAAC TGATTTATAAC TGATTTAAAC TCAATTAACT TACAGCATAA TACAGCATAA TACAGCATAA TACAGCATAA TACAGCATAA TACAGCATAA TACAGCATAA TACAGCATAA TACACATAACT CACCTATAAT CACCTATAAT CACCTATAAT CACCTATAAT CACCTATAAT CACCTATAAT CACCTATAAT	ACTGGCCTTG AGGACAAAGC ATAGATAAGA AGGACAAAGC ATAGATAAGA GCAGCACTCT TTACTGAGTT TTACTGAGTT TTACTACTG GCTTTGCAAG GCAGCACTCT GCTTTGCAAG AGGCTAGAAC CAAATGGCTA AGGCTAGAAC TACAGCACGT AGGGCCATGG AGGCTAGAGA AGGTATATAC TACAGCACGT AGGGCCATGG AGGTTTGCTAA AGTAGGCC TACTTTTCTGT GCAGCACTTATTA AAGTAGGCC CATCTATTTA AAGCCATTTC TGCAGGGTGGG CATCCATCTG TGCAGGGTGG CCATCCATCTG TGCAGGGTGG CCACCACTT AAAATACAAA AAATTCAAA	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT ACAGTTTGG ACACTATGGAT ACAGTTTGG ACACTATGG CAATATATGT ACCTAAGGAT ACCTAAGGAT ACCTAAGGAT CGGTATTAAA ACGTGAGGT ACCTTAAGAA CTGAACAGTA GAGTCTTGGG TTTTAGTATG GATGAAGTGA AGTTAGGAG GTGTGGGCAT AGAGGAACAG AAAAAGGGAA AAAAGGGAA AAAAAGGAA CCTAATTAA CCTGATCTGAT	TOTCAACTIT TCCAGACAT ANAGAAAATT TTCTTCTAT ACTGGCTAT ACTGGCTAT TCTGCAGCCA AATTTGAAA TGCCAGGTAC GATAAAATGT GATGAGTAGG ATTGGAGAAA ATTGAGAGAGA	ATCTCAGCTG TGTTGGAAA ATAAGTGGTG TGGAGCCTGG TTGGAGCCAG TTGGATCTAA TTAAACAGT TTGTTCAGGG GCACTAGGTA CGTTCTCT TTTTGTATGG AAAATGCAGT TGGAGCGGG AAAAGAGGGG CAAAGTTAC AACCTCATA AGCACTTAGAAT AGCATGGATA AGCATGGATA ATCAAGTGC TTAGAAATGAAT AGCATGGATG ATTCAAGTGC TTAGAAATTAC GTTATTCAC ATCGTTCGAT TCCAGGGCCTGT TCCAGGGCCTG TTCAAGGACCT TCCAGGGCCTG TTCAAGGACCT TCCAGGGCCTG TTCAAGGACCT TCCAAGGACCT TCCAAGGACCT TTCAAGGACCC GAAGGAAATTAGG CCACAGGACCT TTCACAGTCC GAAGGAAATT TCAAGGCCC GAAGGAAATT TCAAGGCCC GAAGGAAATT TCACAGTCC GAAGTC GAAGCAT TCACAGTC GAAGTC GAAGTC GAAGCAT TCACAGTC GAAGCAT TCACAGTC GAAGTC TCACAGC GAAGTC GAAGTC GAAGTC GAAGTC GAAGTC GAAGCAT TCACAGTC GAAGCA TCACAGC GAAGTC GAAGC	TGAATCAGA CTCACAAGC AGATGATGA AGATGATGA AAAGGAGTTT TCTGTAAAAGC CTATAAGGGGT CTATAAGGGTT TCTGAGAGGTAT ATGAGGGTAT ATGAGGGTAT ATGAGGATAT TCAGAGGATAT TCAGAGGACAGT TTCAGAGACAGT TTAGACCAGA AGTTCAAATG TATGACCAGA AGTTCAAATG TATGACCAGA ACCAGCATAA ATGATATAAGG TGGATGGATAA ACCAGCTTTTC AACCACCAC AACCAGCATAA ACCAGCTTTTC CTCACTACCAC AGGAACTAC AGGAAATTTCC AGGAAATTTCC GAGAAATTCC GAGAAATTCC GAGAAATTCC GAGAAATTCC GAGAAATTCC GAGAAATTCC AATCAGGACTTC GAGAAATTCC AGCAGCTTCT CAGGAAATTCC AGTAAATCAC AGTAAATCAC AGTAAATCAC AGTAAATCAC AGTAAATCAC AGTAAATCAC AGTAAATCAC AGTAAATCAC AATCAGGACTTC AATCCCAGCTC AATCCCAGCTC AATCCCAGCTC AATCCCAGCTC AATCCCAGCTTC AATCCCAGCTTC AATCCCAGCTTC AATCCCAGCTTC AATCCCAGCTC AATCCCACC AATCCCAGCTC AATCCCACC AATCCCAC	1111 1211 1411 1511 1611 1771 1911 2011 2211 2211 2311 2311 2411 2511 2611 2711 3011 3111 3211 3311 341 3511 3611 3713 3611 4011 4111 4211 4311 4411
CHAMATOL  GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G COTTGEGGA A COCTTCAGGA A COCTTCAGGA A ATAGGAGAMA A ATAGGAGAMA G CTAATAMA G ATATAGACA A CAATATAMA G ATATAGACAA A ATATAAGAA A TATATATOC T GOTCAACACA T TATATTOC T TATATTOC T ATATATAGA A AGAAGAATAA T AGAGGAGTAA A ACAGAAGTAA A ACAGAAGAA A ACAGAACAA A ACAGAACAA A ACAGAACAA A ACAGAACAA A ACAGAACAA A ACAGAACAA A ACACACAACA CCAACACAAC ACTTCAGACA ACT	GLAATAACGA GCAAGGCCATG GCATGAAGGG ACCTGAAGGG ACCTGAAGGG ACCAGATTTAA GCGATTTTAT GCATTTAAAGGG ACAGATTTAA GCAGATTTAA ACTAGAAGGA ACTAGAAGAGG ACAGAAGGAG ACTAGAAGAAGCAT TAGGTAGAAGCAT TAGGTAGAGAAC CTGGCAGAA ACTAGAGAACAT TAGGTAGATA ACTAGAAGCAT TAGGTAGATA CCTGGCAGAA ACTAGAAGCAT ACTAGATAAAGAGCT GAACCATTAAAGAGCT GAACCATTAAAGAGCT GAACCACAATTA CCAGAAGAGCA CTTAAAGAGCT GAACCACAATTA CCAGAGAGAGCA CTTAAAGAGCT GAACACACAG CCTCTCTTAAC ACGTACACAAG GCTCTCTTAACGAGAGCACAAAG CTGGCCAACAA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCCAACA	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA TAGATCATT TATGCATCAA CTCTTTGATC TATGCACTAA CCCATTGATC TAMAGAAGA CCAAMGAAGG TACCACTGAT GACTATTGATC AACTATGAC TAGCACTATA TAGGCTCTTTA AATTAGACTTA ACTAGTAGACTC CCTGGAGACTC CCTGGAGACTC CCTGGAGCCT GGGACCAGGA AAATTATTC AAGTGAGGACT TACACCTGTGA AAATTATTC AAGTGAGGACT TACACCTGTGA TACATTATT TACACCTGGAT TCCATGGAT TCGAGGCTGT AGAGGCTGT AGAGGCTGT AGAGCTGT AGAGGCTGT AGAGGCTGT AGGGCTGCA AGGCTGGCA AGGTGAAATGCTT AGAGGTGAAATGCTT AGAGGTGAAATGCTT AGAGGTGAAATGCTT AGAGGTGAAATGCTT AGAGGTGAAATGCTT AGAGGTGAAATGCTT AGAGGTGAAATGCTT AGAGGTGAAATGCTT AGAGGTGAAATGCTT AGAATAGCTTT AGAATAGCTT AGAATAGCTT AGAATAGCTT AGAATAGCTT AGAATAGCTT AGAAT	TAGAGGGTT TACAGAGGTT TCATTAGGGT TACAGATCCC TAAGGATACCC TAAGGATACACCC TAAGGATACACC AGATGTCCT TAAAGCAGCA ACATTGATAG ATCATCGGG ACATTGATAG ACCATCTGGG GTCAATTA ATTACTCGCI ACATCTGGG GTCAATTAA ATTACTCGT CCCCAACAC TTAGGTCAT TAAGGTCAT AGCCACAGAT TAAGGTCAT AGCCACAGAT TAAGGTCAT ACCCACAGAT TAAGGTCAT ACCCACAGAT TAAGGTCAT ACCCACAGTCAT ACCCACAGTCAT ACTCAGGGAAC CCCAAATTAG TCCAATTAG TGGAGAACCC GAACCTGGGAACCC GAACCTGGGAACCC	CAGAAAGCTE AAACTAGCAA GACATCTGGG TGGGGAAAGG TGGGGAAAGG TGGGGTATAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG ACAGAGATA ACAGATATAA TATGAACCAA TATGAGACAA TATGATATAA TATGAAGGG CACAGTCTGAA TATGGAAGACTCAA ACAGTACT AAACAGTAGT TTTTGTTGAC TGGACATTAA TATTATAAAA TATTATAAAA TATTATAAAA TATTAT	ACTGECCCTE AGTGGTAAA AGGACAAAGC ATAGATAAGA AAGGACAACCT TTATTATTGC TTATTGATTTTTCATCTG GCTTTCCAAG GCAGACTACT AGGCTACAACA AGGCTACAACA AGGCTAGACA AGGCTAGACA ATCCTGAATG TATGATATAC TATGACACACG TATGATATAC TATGACACACG TATGATATAC TACGCACTG TAGGCTAGAG ATGCCTTGTT AGGCCTCGTTAAA AGGCTAGGG AGGTTACCT ACAATACTA TAGGCCATCG AGGTTACAT AGGCCTTTTAAA AGGCATCATG GCAGATCAAG CAACTATTT AAAGCACTT TAGGCCATCT TAGACAACA CAACTATTT AAAGCACATCT TAGACAACA GCAGTCAGCC TAGACTACTA TAGACAACA GCAGTCAGCC TAGACTACTAC TAGACTACTACAACA GCAGTCAGCC TAGACTACTACAACACA GCAGTCAGCC TAGACTACTACACACACACACACACACACACACACACACA	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT CTTCTGCTAT AACAGTTTAG ACAATTACGG CAATTACGG CAATTACGG CAATTACGG CGCTAATGA ACCGTAAGGAT ACCTTAAGGAG CGCTAATGAA ACTTAGGAG TTTTAGTATG GATCAACAGTA TGGTTCACTA AGATTAGGAG AAAAGGGAG AAAAAGGGAG AAAAAGGGAG AATACATTAAC CCTGATTAGAAA CCTGATTAGTTAG CATAGCATTT TATATCAGAG AATACCATT TATATCGGAG ATGGATGCTA AATACCAGG CTAATTTAC CATAGCATTT TATATCGAG ATGGATGCTA AATACCAGG CTAATTGAC CATAGCATTT TATATCGGAG ATGGATGCTA AATACCAGG CACACCATT TCACACGGCCC GACACCACCATT TCACACGGCCC CATTCCCCAC	TOTCAACTIT TCCAGGACAT ANAGAAAATT TCTCTATA ACTGGCTAT ACTGGCTAT TCTGCAGCCA AAATTTGAAA TGGCAGGTAC GATAAAATGT GATAGGTAGG ATTCAGGAG ATTGGACAAT AFGATAGGTAG CAGTACCAGAGAT TTACCTTGCA GATAAAGAT TGGAAAAGT TTGGAAAAGT TTGGAAAAGT TTGCCCAGC CAGTAGAGTAG	ATCTCAGCTC TGTTGGAAA ATAAGTCGTC TGGGGCCTGG TTGTAGCCAG TTGGCTTCTA TTAAAACAGT TGTGTCAGGG GCACTAGGTA TGTGTCAGGG AAAATGCAAT CCGAGGCCCTTAG AACTGGAAT AGCCCTTAG TTAGAAATTA AGCACTGGAAT ATCAAAGTAGG TAAGAAGTAC AATCTGAAAT AGCACTGAAT TAAGGAAGGA TAAAAAAAGAGAT TAAAGAAGAT TCCAAGGACT TCCAAGGACT TCCAAGGACT TCCAAGGACT TCCAAGGACT TCCAAGGCCAA TTCACACTCC GAAGGAAATT TCCACTCCG GAAGGAAATT TCCACTCCC GAAGGAAATT TCCACTCCC GAAGGAAATT TCCACTCCC GAAGGAAATT TCCACTCCC GAAGGAAATT TCCACTCCC GAAGGAAATT TCCACTCCC GGGGCCCTGT CAGCCCTGGCCCTGC CAGCCCTGCCCCCCCCCC	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA GAATAATGA AAAGGAGTTT TCTGTAAAAC TACTAGGGGT TTCAGTAATGA TACCATAGGGTT TCCATATGG ATGGAGGTAT CGTACCACAC AGTTCAATGA TATGGACCAC CTCTAATGA TTCAGGACCAC TTCAGGACCAC TTCAGGACCAC TTCAGGACCAC TTCAGGACCAC TTCAGGACCAC TTTTAGACCAC TTCAGCACAC TTTTAGACCAC AGAAACTTAC AGAAACTTAC AGTAATTAC AGAAACTTAC AGTAAATGAC AGTAAATGAC TCAACCAC TCAACTGAA AGAGCTTCT CAACACAC CTCACTGAA AGAGCTTCT CAACACAC TCAACTGAA AGAAACTTAC CACAGCCTCC CACAGACCAC CTCACCAC CACAAATATG AATCCCAGGCT CACCAGCCAC CACAGCACAA	1111 1211 1411 1511 1611 1711 1911 2011 2211 2211 2211 2311 2411 2511 2611 2711 3011 3011 3011 3011 3011 3011 4011 40
CELAMATOLI AGGACTANA GOTTOTNAGA GOTTOTNAGA GOTTOTNAGA GOTTOTNAGA CCOTTOCAGA ANAGAGNAG ACTATOCACA ANAGAGNAG ACTATOCACA ACT	GAMATAACGA GCAAGGCCAT GGCATCTACC ACCTGAAGGC ACCTGAAGGG ACCAGACTTA GGCATTTAT GGCATTTAT GGCATTTAT GGCATTTAT ACTAGAGGC ACACACTTA ACTAGAGAGC CTAAGAGAG CTAAGAGAG CTAAGAGAG CTAAGAGAG CTAAGAGAC CTAAGAGAG CTAAGAGAC CTAAGAGAC CTAAGAGAC CTAAGAGAC CTAAGAGAC CTAAGAGAC CTAAGAGAC GAGAACC CTAAGAGAC GAGAACC CATAAACTC GATAAACTC GATAACTC GATAAACTC GATAACTC GATAACTC GATAACTC GATAACTC GATACTC GATACT GATCC CTGAGA GCC CATAAACTC GATC GA	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TATTGCATT TATGCATT TATGCATT TATGCATA AATGGGGAAA GCAAAGAAGG TAAACTATGC TACACTGAT GGCTCTTTA TATGGACTA TACACTGAT TACACTGAT TACACTGAT AGTGAGGACC GCGGAGCCG GCGAACCAGA AAATAATACA TTTGGAACTA ACTATGGAC TACACTGTAA ACTATGGAC TACACTGTAA TACACTGCAC TTCATGATT TACACTGCAC AGGTGAAATAC AGGTGAAATAC AGTTGAAGAG AGATGAAAGCT AGTTGAAGAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGAGAG AGTTGAAGAG AGATAGAG AGATAGAG AGGTGAAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGAG AGATAGAG AGATGAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGATAGAG AGATAGA	TAGAGGGT TACAGAGGT TCATAAGGT TCATAAGGT TCATAAGGT TAAGGATAGC TAAGGATAGC AGATGCTGC AGATGCTGC AGATGTCTGTG GCCTGATTTA ATTATCTGCG ACCATTGGG TCATCGGG TCATCGGG TCATCGGG TCATCGGG TCATCGGG TCATCGGG TTAGCTATTA ATTATCTGGG TCTTAATAT TGCTCATTAATAT TGCTCATGGATAAA TCTTAATAT TGGCAGATCC TAGGGAACCC TAGGGAACCC TAGGGAACCC TAGGGAACCC TAGGGAACCC TAGGGAACCC TCAGAGTAAA TCTAACATAT TGGCAGATTAA ACTGAGGAAC TCAGGGAAC TCAGGGAAC TCAGGGAAC TCAGGGAAC TCAGGGAAC TCAGGGAAC TCAGGGAAC TCAGGGAAC TCAGGGAAC TGGGGAACCC TGGAGACCC TGAAACAA TCCAATTAA TCCAATTAA TCCATCAGGAAC TGGGGAACCC TGGAACCC TAACAA TCCAATTAA TCC	CAGAAAGTA GACATATTAGA GACATATTAGA GACATATTAGA GACATATAATA TAGCATTATAGA CATTCAGACA TCAGGCCAGG GAGACATAA ATCAGATATAAT CAAGGCCAGG GAGACATAA TATTGAGACA TATTGAGACA TATTGAGACA TATTGAGACA TATTGAGACA TATTGAGACA TATTGAGAGG CAGAGTACTAA AAACAGTAGT TTTTTTTGAC TGGACATTTA TATTATAAAA CCTGATTGTAG CTGATTGTAG CTGATTGTAG CTGATTGTAG CTCATTGTAG CTCATTGTAG CTCATTGTAG CTCATTGTAG CTCATTGTAG CTCATTGTAG CACTGTAAAT TATCTGGTGG GGGCTAGGAT CACCTATAAT CACCTATAAT CACCTATAAT CACCTATAAT CACCTATAAT GGGGTTACTA GGGTTTACTA GGGTTTACTA GGTTTACTA GGTTTACTA	ACTGECCTTE AGTGAGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAGC TTAATATTGC TTACTGAGTT TITTCATCTG GCTTTGCAAG GAGGTATATA TCCTGCAACA AGGCTAGAAC AGGCTAGAAC ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA AGGATATATAC TACAGCACC TACTGTTAAA CTATTTCTG GCAGTTCATA AGGGCCATCA AGGGCCATCA AGGATCAATA CAATTACCT ACAATACCA TACTGTTAAA CAATTACCT ACAATACCA CAATTACCT ACAATACCA TACTGTTAAA CAATTACCA AAGGACCACT TGCAGGGC CCCAGCACTT AAAATACAAA AAGCAGGCC CCCAGCACTT AAAATACAAA AAGCAGGCC CCCAGCACTT AAAATACAAA AAGCAGGCC CCCAGCACTT AAAATACAAA	TGGTTCCCAC TGGTAGGGAT ATCCAAAGCC ATAAGGATGA GCAATGTGAT GCAATGTGAT ACAGTTTGG ACAACTACGG CAATATATGT ACCTAAGGAG ACCTAAGGAG ACCTAAGGAG CCTAATGTAA AACAGTAGGA TTTTAGTATG GATCAGTAGAT TGGTTCGGGAT TGGACAGTA AGGTTAGGAG AGTTAGGAG CTTAAGGAG CTTAAGGAG ACTTAGGAG AGTTAGGAG TTTCATTAGAAG CCTAATTACA CATAGCATTAA AATATTCAAG CATAGCATTAA AATATTCAAG CATAGCATTAA AATATTCAAG CATAGCATTAA AATATTCAAG CATAGCATT TATAAGGAG AATAGCAGT AATATTCAAG CATAGCATT TATAAGGAG CATAGCATT TATAAGGAG CATAGCATT TATAAGGAG CATAGCATT TATAAGGAG CATAGCATT TATAAGGAG CATAGCATT TATAAGGAG CATAGCATT TATAACGAG CATAGCATT TATAACGAG CATAGCATT TATAACGAG CATAGCATT TATAACGAG CATAGCATC CATTTAGTAG GCACAGCATC TCAGAGGCCC AATTACCTGG GAGATCGCCC AGTGAGCCCCA	TOTCANCTIT TCCAGGCATE ANAGAMANT TTTCTTCTAT ACCTGGCTTC ANACTCTCA ANATTTGANA TGCCAGGTAC GATANANTGG GATANANTGG GATGAGTAGG ATTTCAGAGA ATTGAGAGAT AGATAGGAGTAC GCAGTACAAA GCATTCAGAA TGCCAGGAAGT TTACCTTGCA AGATAGGAT TGGAAAGGT TCGGAAGGT TCGGAAGGT TCGCAGGCAG AGATAATGAAA AGATCAACAA AGATCAACAA TAAAGTCAA ACATCAACAA TAAAGTCAT TAAAGTCT TACTTGCCAGC AATTGTATTTACCAGC ATTTGACAAA TCTAAAGTCT TAAAGTCA TTAAAGTCA TAAAGTCA TTAAAGTCA TTAAAAGTCA TTAAAGTCA TTAAAGTCA TTAAAGTCA TTAAAGTCA TTAAAGTCA TTAAAAGTCA TTAAAGTCA TTAAAAGTCA TTAAAAGTCA TTAAAAGTCA TTAAAAGTCA TTAAAAGTCA TTAAAAGTCA TTAAAAATTAAAAAAAAAA	ATCTCAGGTG TGTAGGAAA ATAAGTGGTG TGGGGCTCTG TTGTAGGCAG TTGAGCAG TTGAGCAG TTGAGCAG TGGGTTCTAG TGTGTCAGGG CCACTAGGTA AGATGCAAT AGATGCAAT AGAAGGGGG CCAAAGTTAG AAAAGAGGGG CCAAAGTTAG AAAAGAGGAA ATCAAAT AGCATGGATA AGCATGGATA TTAAGGAAGT CCTAGCAAGTTA TTAAGGAAGT TTAAGGAAGT TTAAGGATC TCAAGGCCA TCAAGATC TCAAGGCCA TCAAGATC TCAAGGCCAG TCAACGCCC CCAGGCCAGA	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGATTT TCTGTAAAAC TACTAAGGGT TCTAAAGGATT TCTGTAAAC TTCATATGG TTCAATATG TCGATATGG TCGATATGG TATGGAGGTA TGGAGGTA TGGAGGACA AGTTCAAATGT TATGGACTAC CCTCTATGAA TTAGGACTAC CCTCTATGAA ATATATAGG TGGAGGAGGAT ATATATAGG TGGAGATGATA TTTAGAGCAG AAAGAGCTAC ACAACCACAA AGAACCACAC ACAACACACAC	1111 1211 1411 1511 1611 1771 1911 2011 2211 2211 2311 2311 2411 2511 2611 2711 3011 3111 3211 3311 341 3511 3611 3713 3611 4011 4111 4211 4311 4411
CHAMATOL  GOTTOTAGA G GOTTOTAGA G GOTTOTAGA G COTTGAGGA A COCTTGAGGA A COCTTGAGGA A ATGAGATAA A ATGAGATAA A ATGAGATAA A CTAGTATAG G ACTATATAG G ACTATATAG G ACTATATAG G ATTAGACAA G CTAGTATAA A CTAGCACAGA G ATTAGACAA G TOTTGACTA T TATATTTOT T TATATTTOT T TATATTTOTA T TATATTTOTA T TATATTTOTA T TATATTTOTA T TATATTOTA T T TATATTOTA T T T T T T T T T T T T T T T T T T T	GLAATAACGA GCAAGGCCAT GCAAGGCCAT ACCTGAAGGG ACGGGCCAGG FTTGAAAGGG ACAGACTTTAA GGGATTTAT GGCATTTAAGGG ACAGACACTTA ACGAGACACGA ACAGACACTTA ACGAGACACGA ACACACTTA ACGAGACACGA CCATAGAGAA CCAGAGACACATTA ACGAGAACCATTA CCAGAGAACCAT ACGCGAGAA ACGTGGAGAA ACCTGGCAGAA AAGGTCACGA AAGGTCACAG CCACCACGACA AAGGTCACAG CCACCACACAC CCACCACACAC CCACCACACAC CCACCA	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATGCATT TATGCATT TATGCATT TATGCATT TATGCATT TATGCATA CTCTTTGATA GCAAAGAAGG TAAACTAGGC TACACTGAT TACACTGAT TACACTGAT TACATTGAA TTTGGAACTTA TAGATATACA TTTGGAACTTA TAGATATACA TTTGGAACTT TAGATATACA TTTGGACTT TACATTATACA TTTGGAACTT TACATTATACA TTTGGAACTT TACATTATACA TTTGGAACTT TACATTATATT CAGGATGCTGA CTCTTAGAGGA TACACTGTGA GCTTTAATTATT CAGGATGCTGT TTCCATGGATT TGGAGGCTGT TACATTGGAT TAGATGCTGCA AGGTGAAATGC AGGTGAAATGC AGGTGAAATGC AGGTGAAATGC TTTCATTATATT TAGATGGAT TAGATGCTGCA AGGTGAAATGC AGGTGAAATGC AGGTGAAATGC AGTTGAAGAGC AGGTGAAATGC TTTCATTGATGATA AGTTGAAGAGC AGGTGAAATGC AGGTGAAATGC AGTTGAAGAGC AGGTGAAATGC AGTTGAAGAGC AGGTGAAATGC AGTTGAAGAGA ACTCCATAAA	TAGAGGGTT TACAGAGGTT TACAGAGGTT TCATTAGTGT TACAGATCCC TAAGGATAGA TTCATGCTGC AGATGCCTCT TAAGCATCCTT TAAAGCAGCA ATCACTCGTG AGACTCGTC AGACTGGTC ACCACTGCTATTA ATTACTCGCG ACCACTGCTATTA ATTACTCGCG ACCACTGCTATTA ATTACTCGTG GCTCAGTAATA ATTACTACTG GCTCAGTAATA TTACGTCTATTA ATTACATCCCTA ACCACTGCTATA ATTACATCCCTA ACCACTGCTAAACAA TTACGTCATTA ACCACAGAT CCCTAAACAA TTACGTGTAA ACCACAGAT CCCTAAACAA TTACGTTCAC GCCACAGAT TACGCTCAC GCCACAGTTCAC GCCACAGTCAC GCCACAGCTCAC GCACACCCC GAACACCCC GAACACCCCC GAACACCCC GAACACCC GAACACC GAACACCC GAACACC GAACACCC GAACCC G	CAGAAGGTA GACATCTAGGA GACATCTAGGA GACATCTAGGA TGGGGCAAGG AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TGAGGCCAGG GAGACATAA ACAGATACT ATAGGACATAA ATAGGATACT AAAGACTCA ATAGGATACT AAACAGTAGT TTTTGTTGAG CAGATTAGA CCTGATTGAG AAACAGTAGT TTTTGTTGAC TGGACATTAAA CCTGATTATAAA CCTGATTATAA CCTGATTATAA CTGATTAAAC TCAAATAACT TACAGCATAAT TACAGCATAAT TACAGCATAAT TACAGCATAAT TACAGCATAAT CACCTTAATT CACCTTAATT GACTGTAAT CACCTTAATT GACTGTAATT CACCTTAATT GACTGTAATT CACCTTAATT CACCTTAATT GACTGTAATT CACCTTAATT CAC	ACTGECCTTE AGGACALAGC ATAGATAAGA AGGACALAGC ATAGATAAGA AAGGACACTCT TTATTATTAC TTACTGAGTT TTACTGAGTT TTACTGAGTT TTACTGAGTT TTACTCTGAGTT TTACTGAGTA AGGACALAA AGGATCATTA AGGATCATTA ATCTGCALAA AGGATAGACA ATCCTGALAG ATCCTGALAG ATCCTGALAG ATCCTGALAG AGGATCATGT AGGGCCACGG AGGATCATGT AGGACACAA AGTAGGGC ACATTATTA AGGCALAGA CALTTATTA AGGCALAGA AAATACALAA AAATACALAA ATACAAAAA ATACAAAAA ATACAAAAA ATACAAAAA	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT ACAGTTTGGTAT ACAGTTTGG ACAGTTTGG ACAGTTTGG CAATATAGGT GGTCTTAAAA ACGTGAGGT ACCTTAAGGAG CTGTAACAGTA CTGAACAGTA GATTAGTATGGAG GTTTAGTATGG GATCAAGGAG AGTTAGGAG GTGTGGGCAT CCGGAACTAGA ACGTCAACAGTA AACATGAAAG ATAAGGAG ATAATTAAC ACATGAAAGGAA AATATTCGAAC GCAACTAGAT AATATCAAG ACTTAGGAG CCTAATTTAT AATATCAAG CCTAATTTAT AATATCAAG CATTAGCATT TATAACGAAG ATGGATTGAT TATAACGAAG CATTAGCATT TATAACGAAG CATTAGCATT TATAACGAAG CATTAGCATT TATAACGAGG CATATCGTG GCACAGCATT TCAGGCCC AATTAGCTG GAGATGGCAC AATTAGCTGA CAGATTGGCCC AATTAGCTGA CAGATTGGCCC AATTAGCTGA	TOTCHACTIT TCCAGGATA ANAGNANTT TTCTTCTAT ACTGGCTAT ACTGGCTAT TCTGCAGCCA ANTITGANA TGCCAGGTAC GATHANATGT GATGAGTAGG ATTGGACAA ATTGGACAAT AGCATGGACAA ATTGGACAAT TACCTIGCA CAGTTACAGAG TTACCTIGCA GATAATGGATAGGATAGGATAGAGAT TTACCTIGCA TACCTIGCA TACTGGACAAT TACGATAGAGAT TTGGANAGAT TTGGANAGAT TTGGANAGAT TTGGANAGTC AGATCAACAA AGATCAACAA AGATCAACAA ATTTGACAAT TTTAGGAGT TATTGACAAT TTTAGGAGT TATTGACAAT TTTAGGAGT TATTGACAAT TTTAGGAGT ACTTGACAAA TTTAGGAGT TTTAGGAGT ACTTGACAAA TTTAGGAGT TTTAGGAGT ACTTGACCAT ACTTCACAT ACTTGACCAT ACTTGACCAT ACTTGACCAT ACTTCACAT ACTTGACCAT ACTTCACAT ACTTGACCAT ACTTCACAT ACTTCACAT ACTTCACAT ACTTCACAT ACTTCA	ATCTCAGCTG TGTTGGAAA ATAAGTGGTG TGGGGCCTGG TTGGGCTTCTA TTAAACAGT TTGTTCAGGG GCACTAGGTA ATAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAAGAGGGG TTGATGTAT AGCATGGAT ACCATGGAT ATCAAGTGC TTAGAAATTA TAAGGACTGT TTAGAAATTA TAAGGAGTG TTAGAAATTA TAAGGAGTG TTAGAAATTA TAAGGAGTG TTAGAAATTA TAAGGAGTG TTAGAAATTA TAAGGAGTC TTAGAAATTA TAAGGAAGT TTAGAAATTA TAAGGAGTC TTAGAAATTA TAAGGAGTC TTAGAAATTA TAAGGAGTC TTAGAAATTA TAAGGAGTC TTAGAAATTA TAAGGAGTC TTAGAAATTA TCAAGGTC TTAGAAATTA TCAAGGTC TTAGAAATTA TCAAGGTC TTAGAAGTC GAAGGAAATT TCCAGGGCCA TTCACAGTCC GAAGGAAATT ATCACTTTGAG GCGTGCCTGT CAGGCCAGATC CAGGCCAGATC CAGGCCAGATC CAGGCCAGATC CAGGCCAGATC CAGGCCAGC CAGGGCACAGATC CAGGCCAGATC CAGGCCAGATC CAGGCCAGATC CAGGCCAGATC CAGGCCAGATC CAGGCCAGACCAGA	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA GAAATAATGA AAAGGAGTTT TTCTAAAAGC CTACAAAGC CTACAAAGC TTCAAAAGC TTCAAAAGC TTCAAAAGC TTCAAAAGC TTCAAAAGC TTCAAAAGC TTCAAAAAGC TATGAGCACAC AGTGACACACA AGTGACACAC TCAGGAGCACAC TTCAGGAGCACAC TTCAGGAGCACAC TTCAGGAGCACAC TCAGGACACAC AACACACAC AACACCACC TCACCACC TC	1111 1211 1411 1511 1611 1711 1911 2011 2211 2211 2211 2311 2411 2511 2611 2711 3011 3011 3011 3011 3011 3011 4011 40
CELAMATOL A GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G COTTGEGGA A COCTTGEGGA A ATAGGARAM A ATAGGARAM G ATAGGARAM G ATATAGACA G ATATAGAGA A ATATAGGA A ATATAAGGA AGAATAAGT TGAAAAAGGA ACAGGAGTTA G CCAGTAAATA ACAGGAGTTA TGAAAAAGGA ACAGGAGTTA TGTACACT TTTCTCACT TTTCTCACC TTTCTCACC AGATGAATA ACGGACCAGC CCACACCAGC CCACACCACC CCACACCAGC CCACACCACC CCACACCACC CCACACCACC CCACACCAC	GAAATAACGA GCAAGGCCAT GCATGAAGGC ACCTGAAGGG ACCGGCCAGG ACCAGACTTA GGCATTTAT GGCATTTATA GGCATTTATA GGCATTTATA GGCATTTATA ACCGGAAGGC ACCAGAAGGAG ACCAGAAGGAG CTAGAGAGCAT TAGGGAGACAT CCTGGCAGA ACAGAAGCAT TAGGTAGAGAC TAGGGAGCAT TAGGTAGAGAC CTAGAAGCAT CCTGGCAGA ACAGAGCAT TAGATGAGAAGCAT TAGATGAGAAGCAT TAGATGAGAAGCAT TAGATGAGTAGAGAGCAT TAGATGAGAAGCAT CTAGAAAGCAT CATAAACTGAG GACACAAT TTAAAGAGCT GATCCTCTTTA CCTGGCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGGCCAACA CTGGCCAACA CTTCTTTTTTTTTT	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA CGAATTTTC TATGCATT TATGCATT TATGCATT TATGCATCA AATGCGGAAA CCATTGAT GCAAAGAAGG TAACCACTGAT GACTATTGCA TTTGGACT TAGACTCA TTTGGACT TAGACTCA TTTGGACTT AATTAGACT GCGGAACCAGA AATTATCA AGTGAGGCCT GCGGAACCAGA AATTATCA ACTATGAA TTTGAACTC TTCATTAT ACACTTGAA TTCATGAT AGGTGAATGC AGGTGAATGAA AGTTGAAGAC AGGTGAATGC A	TAGAGAGGT TACAGAGGT TCATTAGGTT CATTAGATGC TACAGATCCCC TAAGATCCCC TAAGATCCCC TAAGATACAC AGATGTCCTC TAAAGCAGCA ACATTGATAG ATCATCCGTG GGCTCATTTA ATTATCTGCG GTCATCTGGG GTCACTGC ACAGTGCCTA ATTAGATATAT TAGGTATAATAT TAGGTATAATAT TAGGTATAATAT TAGGTATAATAT TAGGTAAAATTAAGATTAAATAT TAGGTAAAATTAAGATTAAATAT TAGGTATAAAATTAAGATTAAAATTAGGAAAAATTAAGATTAAAATTAGGAAAATTAAGATTAAAATTAGGAAAATTAAGATTAAATATAGGAAAATTAAGATTAAAATTAGGAAAATTAAGATTAAAATTAGGAAAATTAAGTCCAATTAAGATTAAAATTAGGAAAAATTAAGTCTAAAAAAGTCTAAAAAGTCTAAAAAAGTCTAAAAAAAGTCTAA	CAGAMAGCTE AMACTAGCAA GACATCTAGCAA GACATCTAGGAA GACATATACT CAGTATTAAT TGCCTTTTAGG CATTCAGACAA TGAGGCCAGG TGAGGCCAGG GAAGATAAT CGAATGATCTAA ACAGATACT TATTGAGAGAA TATTGAGAGAGA CTAGATCTCAA TATTGAGAGGG CAGAGTCTGAA TATTGGAAGGG CTGGCTACAT TATTTATAAA CCTTATTTTTTTGC TGCCCTAC TGCACATTAA TATTATAAAA CCTTATTATAAC TCAATTAAAC TCAATTAAAC TCAATTAAAC TGCAATAAAT GGCATAAAT GGCTTAATAT CAATGAAGA CAATGAAAA CCAATGAAAAT CCAATGAAAAA	ACTGECCCTE AGTGGTAAA AGGACAAAGC ATAGATAAGA AAGGACAACC TTATTATTGC TTATTATTGC TTATTATTGC TTATTATTGC TTATTATTGC TTATTATTGC GCTTTGCAAG GCAGACTCT TCAGGCACA AGGCTAGACA AGGCTAGACA AGGCTAGACA ATCCTGATAGAC CTATTGTGAG TATGATATAC TACAGCACGT AGGCTCAGGG AGGCTCAGGG AGGCTCAGGG AGGCTTTCCT ACAATAGCCA CAATTTTAA AGGCCATGT GCAGCTGTATAA CAACTATTTA GGATGCAGGC CCACCACTTT GGATGCAGGC CCACCACTCT ACAATACCAAG CCATCCAGGTGG CCCAGCACTT AAAATACAAA GCAGTCAGGC TACAGTGGG AAGGAGGAAGAAAAA GCAGTCAGGCT AAAATACAAA GCAGTCAGGCT AAAGGAAAAAAA CAGAAAAAAA	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTAGT CTTCTGCTAT AACAGTTTAGG ACAACTACGG CAATAATGT ACCTATAGGAT ACCTAGTAGA AACAGTAGGA CGCTAATGAA AACAGTAGGA TTTAGGAG TTTAGGAG AGTTAGGAG AGTTAGGAG AACATGAAA AACATGAAAG AAAAGGGAG AAAAGGGAG AAAAGGGAG AAAAGGGAG TTTCATTATAA CATTAGAAT CATTAGAAT CATTAGATGAA AACATGAAA AACATGAAA CATTAGATGAT TATATCAAG ATGATTCATT TATATCAGG ATGATTCATT TATATCAGG ATGATTCATT TATATCAGAG ATGATTCATT TATATCAGAG ATGATTCATT CACAGGGCCC AACTTAGTAT ACTTATATAAAAA CATTTAGTAG ATGATTCAGT AATTACCAG ATGATTCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAAAAAGAGAAT TCAGAGGCCCC AACTTAGATAAAAAAAGAGAAT TCAGAGGCCCC AACTTAGATAAAAAAAAGAAAAAAAAAA	TOTCAACTIT TCCAGGCATG ANAGAAAATT TTCTTCTAT ACCTGGCTTT TCTGCAGCCA AAATTTGAAA TGGCAGGTAC GATAAAATGT GATGAGTAGG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG CAGTTACAGAGT TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCAGC CAGTAAACAA GAGTAGAGAT TTGCCAGC CAGTAGACAA TAAATGTGAT AAAAAAAGGA TTTGACAAA TAAATGTGAT AAAAAAATGTGAT AAAAAAATGTGAT AAAAAAATGTGAT AAAAAAAA	ATCTCAGCTC TCTTCGAAAA ATAAGTCCTG TGGAGCCTCG TTGTAGCCAG TTGGATCTAGCCAG TGGATTCTA TTAAAACAGT TGTGTCAGGG CCACTAGGTA AGAGGGCGCC AAAAGAGGGC AAAAGAGGGC AAACAGGATAC AACCTTAGAAT AGCATTGAAT AGCATTGAAT AGCATGGATA TTAAGGAAGT TTAAGGAAGT TTAAGGAAGT TTAAGGAAGT TTAAGGATCA TTAAGGAAGT TTAAGGATCA TTAAGGATCA TTAAGGATCA TTAAGGATCA TTCACAGTCC TCACAGTCC TTCACAGTCC TCACAGTCC TCACAGTC TCACA	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA AAAGGATTT TCTGTAAAAC TACTAGGGGT TACTAGGGGT TCCATATGG ATGAGGGTAT TCCATATGG TACCATATGG ATGAGGGTAT TCCATATGG TATGAGCACAGA AGTTCAAATGT TATGACACAGA TATGAGCACAGT TCAGGAGCACA TTATGAGCACA TTATGAGCACA TTATATAAAGG TCGATGGATAAA ATATATAAGG AAAAGAGCTAC TCAGCACAC TCTACTGAA AGAACCTAGA AGAAACCTAGA AGAAACCTAGA AGAAACCTAGA AGAAACCTAGA TGATAATTGAC AGTAAATGAG TCAGCACAC TCACCACAC TCACCAC TCACCACAC TCACCAC TC	1111 1211 1311 1411 1611 1711 1811 1911 2011 2011 2011 2011 2011 2011 3011 30
CELAMATOL A GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G COTTGEGGA A COCTTGEGGA A ATAGGARAM A ATAGGARAM G ATAGGARAM G ATATAGACA G ATATAGAGA A ATATAGGA A ATATAAGGA AGAATAAGT TGAAAAAGGA ACAGGAGTTA G CCAGTAAATA ACAGGAGTTA TGAAAAAGGA ACAGGAGTTA TGTACACT TTTCTCACT TTTCTCACC TTTCTCACC AGATGAATA ACGGACCAGC CCACACCAGC CCACACCACC CCACACCAGC CCACACCACC CCACACCACC CCACACCACC CCACACCAC	GAAATAACGA GCAAGGCCAT GCATGAAGGC ACCTGAAGGG ACCGGCCAGG ACCAGACTTA GGCATTTAT GGCATTTATA GGCATTTATA GGCATTTATA GGCATTTATA ACCGGAAGGC ACCAGAAGGAG ACCAGAAGGAG CTAGAGAGCAT TAGGGAGACAT CCTGGCAGA ACAGAAGCAT TAGGTAGAGAC TAGGGAGCAT TAGGTAGAGAC CTAGAAGCAT CCTGGCAGA ACAGAGCAT TAGATGAGAAGCAT TAGATGAGAAGCAT TAGATGAGAAGCAT TAGATGAGTAGAGAGCAT TAGATGAGAAGCAT CTAGAAAGCAT CATAAACTGAG GACACAAT TTAAAGAGCT GATCCTCTTTA CCTGGCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGCCAACA CTGGGCCAACA CTGGCCAACA CTTCTTTTTTTTTT	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA CGAATTTTC TATGCATT TATGCATT TATGCATT TATGCATCA AATGCGGAAA CCATTGAT GCAAAGAAGG TAACCACTGAT GACTATTGCA TTTGGACT TAGACTCA TTTGGACT TAGACTCA TTTGGACTT AATTAGACT GCGGAACCAGA AATTATCA AGTGAGGCCT GCGGAACCAGA AATTATCA ACTATGAA TTTGAACTC TTCATTAT ACACTTGAA TTCATGAT AGGTGAATGC AGGTGAATGAA AGTTGAAGAC AGGTGAATGC A	TAGAGAGGT TACAGAGGT TCATTAGGTT CATTAGATGC TACAGATCCCC TAAGATCCCC TAAGATCCCC TAAGATACAC AGATGTCCTC TAAAGCAGCA ACATTGATAG ATCATCCGTG GGCTCATTTA ATTATCTGCG GTCATCTGGG GTCACTGC ACAGTGCCTA ATTAGATATAT TAGGTATAATAT TAGGTATAATAT TAGGTATAATAT TAGGTATAATAT TAGGTAAAATTAAGATTAAATAT TAGGTAAAATTAAGATTAAATAT TAGGTATAAAATTAAGATTAAAATTAGGAAAAATTAAGATTAAAATTAGGAAAATTAAGATTAAAATTAGGAAAATTAAGATTAAATATAGGAAAATTAAGATTAAAATTAGGAAAATTAAGATTAAAATTAGGAAAATTAAGTCCAATTAAGATTAAAATTAGGAAAAATTAAGTCTAAAAAAGTCTAAAAAGTCTAAAAAAGTCTAAAAAAAGTCTAA	CAGAMAGCTE AMACTAGCAA GACATCTAGCAA GACATCTAGGAA GACATATACT CAGTATTAAT TGCCTTTTAGG CATTCAGACAA TGAGGCCAGG TGAGGCCAGG GAAGATAAT CGAATGATCTAA ACAGATACT TATTGAGAGAA TATTGAGAGAGA CTAGATCTCAA TATTGAGAGGG CAGAGTCTGAA TATTGGAAGGG CTGGCTACAT TATTTATAAA CCTTATTTTTTTGC TGCCCTAC TGCACATTAA TATTATAAAA CCTTATTATAAC TCAATTAAAC TCAATTAAAC TCAATTAAAC TGCAATAAAT GGCATAAAT GGCTTAATAT CAATGAAGA CAATGAAAA CCAATGAAAAT CCAATGAAAAA	ACTGECCCTE AGTGGTAAA AGGACAAAGC ATAGATAAGA AAGGACAACC TTATTATTGC TTATTATTGC TTATTATTGC TTATTATTGC TTATTATTGC TTATTATTGC GCTTTGCAAG GCAGACTCT TCAGGCACA AGGCTAGACA AGGCTAGACA AGGCTAGACA ATCCTGATAGAC CTATTGTGAG TATGATATAC TACAGCACGT AGGCTCAGGG AGGCTCAGGG AGGCTCAGGG AGGCTTTCCT ACAATAGCCA CAATTTTAA AGGCCATGT GCAGCTGTATAA CAACTATTTA GGATGCAGGC CCACCACTTT GGATGCAGGC CCACCACTCT ACAATACCAAG CCATCCAGGTGG CCCAGCACTT AAAATACAAA GCAGTCAGGC TACAGTGGG AAGGAGGAAGAAAAA GCAGTCAGGCT AAAATACAAA GCAGTCAGGCT AAAGGAAAAAAA CAGAAAAAAA	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTAGT CTTCTGCTAT AACAGTTTAGG ACAACTACGG CAATAATGT ACCTATAGGAT ACCTAGTAGA AACAGTAGGA CGCTAATGAA AACAGTAGGA TTTAGGAG TTTAGGAG AGTTAGGAG AGTTAGGAG AACATGAAA AACATGAAAG AAAAGGGAG AAAAGGGAG AAAAGGGAG AAAAGGGAG TTTCATTATAA CATTAGAAT CATTAGAAT CATTAGATGAA AACATGAAA AACATGAAA CATTAGATGAT TATATCAAG ATGATTCATT TATATCAGG ATGATTCATT TATATCAGG ATGATTCATT TATATCAGAG ATGATTCATT TATATCAGAG ATGATTCATT CACAGGGCCC AACTTAGTAT ACTTATATAAAAA CATTTAGTAG ATGATTCAGT AATTACCAG ATGATTCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAG ATGATCAGAAAAAAGAGAAT TCAGAGGCCCC AACTTAGATAAAAAAAGAGAAT TCAGAGGCCCC AACTTAGATAAAAAAAAGAAAAAAAAAA	TOTCAACTIT TCCAGGCATG ANAGAAAATT TTCTTCTAT ACCTGGCTTT TCTGCAGCCA AAATTTGAAA TGGCAGGTAC GATAAAATGT GATGAGTAGG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG CAGTTACAGAGT TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCAGC CAGTAAACAA GAGTAGAGAT TTGCCAGC CAGTAGACAA TAAATGTGAT AAAAAAAGGA TTTGACAAA TAAATGTGAT AAAAAAATGTGAT AAAAAAATGTGAT AAAAAAATGTGAT AAAAAAAA	ATCTCAGCTC TCTTCGAAAA ATAAGTCCTG TGGAGCCTCG TTGTAGCCAG TTGGATCTAGCCAG TGGATTCTA TTAAAACAGT TGTGTCAGGG CCACTAGGTA AGAGGGCGCC AAAAGAGGGC AAAAGAGGGC AAACAGGATAC AACCTTAGAAT AGCATTGAAT AGCATTGAAT AGCATGGATA TTAAGGAAGT TTAAGGAAGT TTAAGGAAGT TTAAGGAAGT TTAAGGATCA TTAAGGAAGT TTAAGGATCA TTAAGGATCA TTAAGGATCA TTAAGGATCA TTCACAGTCC TCACAGTCC TTCACAGTCC TCACAGTCC TCACAGTC TCACA	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA AAAGGATTT TCTGTAAAAC TACTAGGGGT TACTAGGGGT TCCATATGG ATGAGGGTAT TCCATATGG TACCATATGG ATGAGGGTAT TCCATATGG TATGAGCACAGA AGTTCAAATGT TATGACACAGA TATGAGCACAGT TCAGGAGCACA TTATGAGCACA TTATGAGCACA TTATATAAAGG TCGATGGATAAA ATATATAAGG AAAAGAGCTAC TCAGCACAC TCTACTGAA AGAACCTAGA AGAAACCTAGA AGAAACCTAGA AGAAACCTAGA AGAAACCTAGA TGATAATTGAC AGTAAATGAG TCAGCACAC TCACCACAC TCACCAC TCACCACAC TCACCAC TC	1111 1211 1311 1411 1611 1711 1811 1911 2011 2011 2011 2011 2011 2011 3011 30
COLAMATOL A  GOTTOTAMA G  GOTTOTAMA G  GOTTOTAMA G  COTTGAGGA A  COCTTGAGGA A  ACCOTTGAGGA A  ATGGGATAA A  ATGGGATAA A  ATAGGAAAA G  ATTAGAGAAA G  CAAGTATAA G  CAAGTATAA G  CAAGTATAA G  TATTAACAA G  TATTATTOC T  TATTATTOC T  TATATTTOC T  TATATTTOC T  TATATTOC T  TATATTTOA T  TATATTTOA T  TATATTTOA T  TATATTTOA T  TATATTOA T  TATATTTOA T  CAACACAGG  CAACAACACAG  CAACACAACACACAC	GAAATAACGA GCAAGGCCAT GCAAGGCCATGA GCAGGCCAGGC	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATGCATT TATGCATT TATGCATT TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC TACACTGAT GACTATTGCA TACACTGAT AGGATGCAT AGGAACAC AGGATGCATC AGGAACAC AGGATGCATC AGGAACAC AGGACAC AGGAACAC AGGAC AGGAACAC AGGAACAC AGGAAC AGGAACA	TAGAGAGGTT TACAGAGGTT TCATTAGGTT TCATTAGTGT TAGAGTTCCC TAAGGATAGA TTCATGCTGC AGATGCCTCT TAAAGCAGCA ATCACTCGTG AGACCTGGTC AGACCTGGTC ACCACTGCTA ATTACTCGCG ACCACTGCTA ATTACTCGCG ACCACTGCTAAACAA ATTACTCGTG ACCACTGCTAAACAA TTAGGTCATTAAATAT GGGAGAATCTT AGACAGCACACAT TTAGGTCATTAAATAT GGCAGATACAT TAGGTCATTAAATAT GGCAGATACAT TAGGTTATAA ATTAGGTCATT AGCCACAGAT TTAGGTCATT AGCCACAGAT TTAGGTCATT AGCCACAGAT TCCAATTTAG GTGAAACCC GAACCTGGGA AAAAGTCTA GTGAACACC GAACCTGGGA CTGCAATACCC GAACCTGGGA AAAAGTCTA GTGAACACC GGAACTAGCA CTCAAGCAT CCAATTTAGGTCATACAA TCCAATTTAGGTCACAC GTGAACACCC GAACACCC GAACACCC GAACACCC GAACACCC GGAACTAGTCCTC GGAACACCC GGAACTAGCT TCCAGCCACAC TCCAGCCCACC TCCAGCCACC TCCAGCCCACC TCCAGCCACC TCCAGCCACC TCCAGCCCACC TCCAGCCCACC TCCAGCCACC	CAGAAGGT GACATCTAGGA GACATCTAGGA GACATCTAGGA GACATCTAGGA TGGGGCAAGGA TGAGGCCAGG GAGACATCA ATCAGATATA TCATCTAGACA ATCAGATATA TATCAGACA ATCAGATATA TATCAGACA TATGGAAGGG CAGAGTCTGA AACAGTAGT TTTTTTGTTGA AACAGTAGT TTTTTTTTTT	ACTGGCCTTG AGGACAAAGC ATAGATAAGA AGGAGAAAGC ATAGATAAGG GCAGCACTCT TTACTGAGTT TTACTGGGTT TTACTACTG GCTTTGCAAG GCAGCACATA AGGGCAACA AGGCTAGAAC CAAATAGCAA AGGCTAGAC TACAGCACGT AGGGCCATGG AGGGCCATGG AGGGCCATGG AGAGTACAC CAATAGTAC CAATAGTAC CAATAGTAC CAACTATTA AGGCAAGGC CAACCACTC TACAGCACGT AAATAGCAA CAACTATTA CAACTATTA CAACTATTA AGGCAAGC CCCAGCCACT AAAATACAA CAACTATTA CAACTATTA AGGCAAGC CCCAGCCACT AAAATACAA CAGAAAAAAAA CAGAAAAAAAA CAGAAAAAAAA	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT ACAGTTTGGTAT ACAGTTTGG CAATATATGT ACAGTTAGA ACAGTTAGA ACAGTTAGA ACAGTTAGA ACGTCAGGG ACCTTAGGAG CGTAATGAA CTGAACAGTA GAGTCTTGGG TTTTAGTATG GATCAAGGAG AGTTAGGAG TGGTCAGGA ACAGTAGA ACAGTAGAA CCGAATTAGA CCGAATTAGA CCGAATTAGA CCGAATTAGA ACATGAAGGA AAAAGGGA ATAATTACAGA ATTAATCAGA ATTAATCAGA ATTAATCAGA ATGATTAGT TATAATCAGA ATGATTAGT GCACAGCATT TATAATCAGA ATGATTAGT GCACAGCATT TATAATCAGA ATGATTCGT GCACAGCATT TATAATCAGA ATGATTCGT GCACAGCATT TATAATCAGA ATGATTCGT GCACAGCATT TATAATCAGA ATGATTCGT GCACAGCATT TATATCAGA ATTAATCAGA ATGATTCGT ACATTAGT ACATTAGT ACATTAGT ATTTATACA ATTTACA ATTTATACA ATTTAT	TOTCAACTIT TCCAGACATE ANAGANANTI TTCTTCTAT ACCTGGCTAT ACCTGGCTAT TCTGCAGCCA ANTITGAAA TGGCAGGTAC GATANANTGT GATAGGTAGG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTGGATAGG GCAGTACAAA GACTTTGGT TTACCTTGGA GTAATGGAT GGATAGGTAG CTGTGGAAAGAT TTGGCAAAGAT TTGGAAAAGAT TTGGAAAAGAT TTGGAAAGAT TTGGAAAGAT TTGGAAAGAT AGTTAGCAGGAG CAGTTAGTCAA GATTAGGATAG AGATTAGCAT AAAAAATGGA TTTAGAGAT TTAAGAGTT ACTTGACAA TTTAGGAGTAG CAGTTAGTAT ACTTGACAA TTTAGGAGT TTAAGAGT TTAAGAGT AAAAAAATGGA TTTAAGAGT TTAAGAGT AAAAAATGGA CACTGCAGC CACTGCACC AAAGGAACTT CACTCCCGTT CCCTTGGTCC TACACACACAC CATTCCCCTT TACACACACAC CATTCCCCTT TACACACAC	ATCTCAGCTG TGTTGGAMA ATAMCTCGTG TGTGGGACCAGG TTGGGCTTCTA TTAMACAGT TTGTTCAGGG GCACTAGGTA TTTTTGTATCG AMATGCAT TTTTGTATCG AMATGCAT TTTTGTATCG AMATGCAT TTGTATCGA AMATGCAT TCATGCAT AMACGCCTTAG CAMAGTTAC CAMAGTTAC CTATGCAGT ATTCAAGTTC TTAGAATTAC TCAGGGCCC TAGAGGAATT TCCAGGGCCC TAGAGGAATT TCCAGGGCCC TTAGAGGACC TTAGAGGACC TCTACCTTGGG CCAGGCCCTGT CCAGGCCCTGT CCAGGCCCTGT CAGCCCTGGC CCAGGCCAGT TTTTTCAGATC TCTTTCAGATC TTTTTCAGATC TTTTTCAGATC TTTTTCAGATC TATTTCATTT	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA GAATAATGA AAAGGACTT TCTGTAAAAGC CTTTAAATGT TTCATATGAGGTT TTCATATGAGGTT TTCATATGAGGTT TATGAGGTAT AGTTCAAATG TATGAGGTAT TATGAGGACA AGTTCAAATG TATGAGCACA AGTTCAAATG TATGACACACA AGTTCAAATG TCAGGAGCACA TTTAGACCAC TCAGGACTATA ACCAGCATA ACCAGCATA ACCAGCATA ACCAGCATA ACCAGCATA TCAGACACAC CTCTACTGAA ACCAGCATA TCAGCACACA CTCTACTGAC AGGAATATTGC AGGAAATTGG TCAGCACAC ACCAGCATA ACCAGCATA TCAGCAGCAT AATCCAGCT AATCCAGCT TCAGCAGATA TTTCAGCAGC TCAGCAGATA TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TATTCAGCAGT TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAAA TACATGAGAAA TACATGAGAAA TACATGAAAA TACATGAAAAA TACATGAAAAA TACATGAAAAA TCCAGCTACACAACA TACATGAAAAA TACATGAAAAA TACATGAAAAA TCCAGCATACAACAACAACAACAACAACAACAACAACAACAACAA	1111 1211 1411 1511 1611 1711 1911 2011 2211 2211 2211 2311 2411 2511 2611 2711 3011 3111 3111 3411 3511 3611 4011 4411 4411 4411 4511 4611 4711 4611 5011
COLAMATOL A  GOTTOTAMA G  GOTTOTAMA G  GOTTOTAMA G  COTTGAGGA A  COCTTGAGGA A  ACCOTTGAGGA A  ATGGGATAA A  ATGGGATAA A  ATAGGAAAA G  ATTAGAGAAA G  CAAGTATAA G  CAAGTATAA G  CAAGTATAA G  TATTAACAA G  TATTATTOC T  TATTATTOC T  TATATTTOC T  TATATTTOC T  TATATTOC T  TATATTTOA T  TATATTTOA T  TATATTTOA T  TATATTTOA T  TATATTOA T  TATATTTOA T  CAACACAGG  CAACAACACAG  CAACACAACACACAC	GAAATAACGA GCAAGGCCAT GCAAGGCCATGA GCAGGCCAGGC	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATGCATT TATGCATT TATGCATT TATGCATC TATGCATC TATGCATC TATGCATC TATGCATC TACACTGAT GACTATTGCA TACACTGAT AGGATGCAT TACACTGAT AGGATGCAT AGGAACAC AGGATGCATC AGGAACAC AGGATGCATC AGGAACAC AGGACAC AGGAACAC AGGAC AGGAACAC AGGAACAC AGGAAC AGGAACA	TAGAGAGGTT TACAGAGGTT TCATTAGGTT TCATTAGTGT TAGAGTTCCC TAAGGATAGA TTCATGCTGC AGATGCCTCT TAAAGCAGCA ATCACTCGTG AGACCTGGTC AGACCTGGTC ACCACTGCTA ATTACTCGCG ACCACTGCTA ATTACTCGCG ACCACTGCTAAACAA ATTACTCGTG ACCACTGCTAAACAA TTAGGTCATTAAATAT GGGAGAATCTT AGACAGCACACAT TTAGGTCATTAAATAT GGCAGATACAT TAGGTCATTAAATAT GGCAGATACAT TAGGTTATAA ATTAGGTCATT AGCCACAGAT TTAGGTCATT AGCCACAGAT TTAGGTCATT AGCCACAGAT TCCAATTTAG GTGAAACCC GAACCTGGGA AAAAGTCTA GTGAACACC GAACCTGGGA CTGCAATACCC GAACCTGGGA AAAAGTCTA GTGAACACC GGAACTAGCA CTCAAGCAT CCAATTTAGGTCATACAA TCCAATTTAGGTCACAC GTGAACACCC GAACACCC GAACACCC GAACACCC GAACACCC GGAACTAGTCCTC GGAACACCC GGAACTAGCT TCCAGCCACAC TCCAGCCCACC TCCAGCCACC TCCAGCCCACC TCCAGCCACC TCCAGCCACC TCCAGCCCACC TCCAGCCCACC TCCAGCCACC	CAGAAGGT GACATCTAGGA GACATCTAGGA GACATCTAGGA GACATCTAGGA TGGGGCAAGGA TGAGGCCAGG GAGACATCA ATCAGATATA TCATCTAGACA ATCAGATATA TATCAGACA ATCAGATATA TATCAGACA TATGGAAGGG CAGAGTCTGA AACAGTAGT TTTTTTGTTGA AACAGTAGT TTTTTTTTTT	ACTGGCCTTG AGGACAAAGC ATAGATAAGA AGGAGAAAGC ATAGATAAGG GCAGCACTCT TTACTGAGTT TTACTGGGTT TTACTACTG GCTTTGCAAG GCAGCACATA AGGGCAACA AGGCTAGAAC CAAATAGCAA AGGCTAGAC TACAGCACGT AGGGCCATGG AGGGCCATGG AGGGCCATGG AGAGTACAC CAATAGTAC CAATAGTAC CAATAGTAC CAACTATTA AGGCAAGGC CAACCACTC TACAGCACGT AAATAGCAA CAACTATTA CAACTATTA CAACTATTA AGGCAAGC CCCAGCCACT AAAATACAA CAACTATTA CAACTATTA AGGCAAGC CCCAGCCACT AAAATACAA CAGAAAAAAAA CAGAAAAAAAA CAGAAAAAAAA	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT ACAGTTTGGTAT ACAGTTTGG CAATATATGT ACAGTTAGA ACAGTTAGA ACAGTTAGA ACAGTTAGA ACGTCAGGG ACCTTAGGAG CGTAATGAA CTGAACAGTA GAGTCTTGGG TTTTAGTATG GATCAAGGAG AGTTAGGAG TGGTCAGGA ACAGTAGA ACAGTAGAA CCGAATTAGA CCGAATTAGA CCGAATTAGA CCGAATTAGA ACATGAAGGA AAAAGGGA ATAATTACAGA ATTAATCAGA ATTAATCAGA ATTAATCAGA ATGATTAGT TATAATCAGA ATGATTAGT GCACAGCATT TATAATCAGA ATGATTAGT GCACAGCATT TATAATCAGA ATGATTCGT GCACAGCATT TATAATCAGA ATGATTCGT GCACAGCATT TATAATCAGA ATGATTCGT GCACAGCATT TATAATCAGA ATGATTCGT GCACAGCATT TATATCAGA ATTAATCAGA ATGATTCGT ACATTAGT ACATTAGT ACATTAGT ATTTATACA ATTTACA ATTTATACA ATTTAT	TOTCAACTIT TCCAGACATE ANAGANANTI TTCTTCTAT ACCTGGCTAT ACCTGGCTAT TCTGCAGCCA ANTITGAAA TGGCAGGTAC GATANANTGT GATAGGTAGG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTCAGGAG ATTGGATAGG GCAGTACAAA GACTTTGGT TTACCTTGGA GTAATGGAT GGATAGGTAG CTGTGGAAAGAT TTGGCAAAGAT TTGGAAAAGAT TTGGAAAAGAT TTGGAAAGAT TTGGAAAGAT TTGGAAAGAT AGTTAGCAGGAG CAGTTAGTCAA GATTAGGATAG AGATTAGCAT AAAAAATGGA TTTAGAGAT TTAAGAGTT ACTTGACAA TTTAGGAGTAG CAGTTAGTAT ACTTGACAA TTTAGGAGT TTAAGAGT TTAAGAGT AAAAAAATGGA TTTAAGAGT TTAAGAGT AAAAAATGGA CACTGCAGC CACTGCACC AAAGGAACTT CACTCCCGTT CCCTTGGTCC TACACACACAC CATTCCCCTT TACACACACAC CATTCCCCTT TACACACAC	ATCTCAGCTG TGTTGGAMA ATAMCTCGTG TGTGGGACCAGG TTGGGCTTCTA TTAMACAGT TTGTTCAGGG GCACTAGGTA TTTTTGTATCG AMATGCAT TTTTGTATCG AMATGCAT TTTTGTATCG AMATGCAT TTGTATCGA AMATGCAT TCATGCAT AMACGCCTTAG CAMAGTTAC CAMAGTTAC CTATGCAGT ATTCAAGTTC TTAGAATTAC TCAGGGCCC TAGAGGAATT TCCAGGGCCC TAGAGGAATT TCCAGGGCCC TTAGAGGACC TTAGAGGACC TCTACCTTGGG CCAGGCCCTGT CCAGGCCCTGT CCAGGCCCTGT CAGCCCTGGC CCAGGCCAGT TTTTTCAGATC TCTTTCAGATC TTTTTCAGATC TTTTTCAGATC TTTTTCAGATC TATTTCATTT	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA GAATAATGA AAAGGACTT TCTGTAAAAGC CTTTAAATGT TTCATATGAGGTT TTCATATGAGGTT TTCATATGAGGTT TATGAGGTAT AGTTCAAATG TATGAGGTAT TATGAGGACA AGTTCAAATG TATGAGCACA AGTTCAAATG TATGACACACA AGTTCAAATG TCAGGAGCACA TTTAGACCAC TCAGGACTATA ACCAGCATA ACCAGCATA ACCAGCATA ACCAGCATA ACCAGCATA TCAGACACAC CTCTACTGAA ACCAGCATA TCAGCACACA CTCTACTGAC AGGAATATTGC AGGAAATTGG TCAGCACAC ACCAGCATA ACCAGCATA TCAGCAGCAT AATCCAGCT AATCCAGCT TCAGCAGATA TTTCAGCAGC TCAGCAGATA TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TCAGCAGCT TATTCAGCAGT TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAA TACATGAGAAA TACATGAGAAA TACATGAGAAA TACATGAAAA TACATGAAAAA TACATGAAAAA TACATGAAAAA TCCAGCTACACAACA TACATGAAAAA TACATGAAAAA TACATGAAAAA TCCAGCATACAACAACAACAACAACAACAACAACAACAACAACAA	1111 1211 1411 1511 1611 1711 1911 2011 2211 2211 2211 2311 2411 2511 2611 2711 3011 3111 3111 3411 3511 3611 4011 4411 4411 4411 4511 4611 4711 4611 5011
CHAMATOLI  AGGACTANAA  GOTTOTNAGA  GOTTOTNAGA  GOTTOTNAGA  CCOTTGAGGA  ACCOTTGAGGA  ATGAGATAA  ATGAGATAA  ATGAGATAA  ATGAGACAA  CTAGTATAAG  ACTATATAAG  ACTATATAAG  ATGAGACAAC  ATTAGACAA  TATTAGACAA  TATTAGACAA  TATTAGACAA  TATTAGACAA  TATTAGACAA  TATTAGACAA  ACAGAACAAC  TATTACTTOL  TATACTTOL  TATACTTOL  TATACTTOL  TATACTTOL  TATACTTOL  TATACTTOL  TATACTTOL  TATACTTOL  TATACTTOL  TATACTCACT  CAGTAAAAA  ACAGAACTTA  TATCTCACT  TATCTCACACA  GOTAACACA  GOTAACACA  GOTAACACA  TATCTCACACA  CACACACACA  ACTTCGCAGC  CACTTCACACA  CACACACACA  AUTAGACTOT  TATCTCTTCACACA  TATCTCACACA  ACACACAC	GLANTACCA GCAAGGCCAT GCAAGGCCAG GCCTGAAGGG AGGGGCCAGG AGGGGCCAGG ACCAGACTTAA GGCATTTAT GGCATTTAT GCATTAAGGG ACAGACTCT AAGAGAC CTAAGGAAC CTAAGGAAC CTAAGGAAC CTAAGGAAC CTAAGGAAC CTAAGGAAC CTAAGGAAC CTAAGGAAC CTAAGGAAC CTAAGAAC CTAGAGAAC CTAGAGAAC CTAGAGAAC CTAGAGAAC CTACCCCAGAGA AAGACAAT GAGCACAAT GAGCACAAT GAGCACAAT CATGGCCAACA CTGGCCAACA CTGAGGACAC CTGAGAAC CACTTTAAGA CACTTTAACAC ACACTTTAACAC ACACTTTAACAC ACACTTAACAC CACTTTAACAC CA	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTTC TATGCATT TATGCATT TATGCATT TATGCATT TATGCATT TATGCATT TATGCATT TATGCATT TATGCATA GCAAAGAAGA GCAAAGAAGA TACACTTGAA TACACTTTGAA TACATATACA TATGGACTTA AAATAAGATT AGTGAGGACCT TACATATACA TATGGACTTA AAGTAATACTCT AAAGTAATTCTC TACATGAGA TACACTGTGA GCTTGAAGATT ACACTGTGA CTCTAGAGGA TACACTGTGA ACACTGTGA AGGAGCTT TACATGATT AGATGGATTATT TACATGAT AGTTGAAAA AGGAACAA AGGAACAA AAGGAACAA AAGGAACAA AAGGAACAA AAGGAACAA AAGGAACAA AAGGAACAA AAGGAACAA AAGGAACAA AAGGAACAA AACACTAAAA AAGGAACAA AACACCTAAAAA AAGGAACAA AACACCTAAAAA AAGGAACAAC AACACTAAAAA AAGGAACAAC AACACTAACAA AACACCATAAAA AACACCATAAAA AACACCATAAAA AACACCATAAAA AACACCATAAAA AACACCATAAAA AACACCATAAAA AACACCCATAAAA AACACCCATAAAA AACACCCAAAAA AACACCCAAAAA AACACCCAAAAA AACACCCAAAAA AACACCCAAAAAA	TAGAGAGGTT TACAGAGGTT TCATAAGGTT TCATAAGGTT TAGAGATAGCA TAGAGATAGCA TAGAGATAGCA AGATGTCCTT TAAAGCATAGCA AGATGTCCTT TAAAGCAGCA ATTACTGGG GCTGATTTA ATTACTGGG GTCAGTACTA ATTACTGGG GTCAGTAATAG ATTAGATAGT TAGAGTAATAT GCGCAGATACT TAGAGTATAA TTAGGTCATT GAGAAATTA GGCAGATACA TTAGGTCAT TAGGTCATTA TTAGGTCATTA TTAGGTCATTA TTAGGTCATTA GCCAAGAT CCCTAAACAA TTAGGTCAT TAGGTCAT GAGAAATTAG TGCAATTAG TGGAAACCC GAACCTGGGA  AAAAGTCTA GTGAAACCC GGAAGTAGCT TCCAGGCAAC TTAGGTCAT TTAGGTTAA TTAGGTTAA TTAGGTTAA TTAGGTTAA TTAGGTTAA TTAGGTTAA TTAGGTTAA TTAGGTTAA	CAGAAAGETA GACATCTAGGA GACATCTAGGA GACATCTAGGA TGGGGCAAAGA CAATCAGACA TGAGGCCAGG GAGACATAA ATCAGATAAT ATCAGATAAT ATCAGATAAT ATCAGATAAT ATCAGATAAT ATCAGATAAT ATCAGATAAT ATCAGATAAT ATCAGATAAT TTTTTTTGTGAC CAGATTATAAC TTATTATAAC TTATTATAAC TTATTATAAC TCAATTAAAT TCAATCAA	ACTGECCTTE AGGACALAGC ATAGATAAGA AGGACALAGC ATAGATAAGA AAGGACACACT TTACTGAGTT TGAGAGGTTT TGAGAGGTTT TGAGAGGTTT TGAGAGGTTT TGAGAGGTTATACACACACACACACACACACACACACACA	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT ACAGTTTCGCTAT ACAGTTTCG ACACTATCGG CAATATAGGT ACCTTAGGAG ACCTTAGGAG ACCTTAGGAG CCGTAATGAT ACCTTAGGAG TTTAGTATG GATCAGGAG TTTAGTATG CCGAACAGA TCGGTCGGCAT ACGTTCAGCA TGGTTCAGCAT ACGTTCAGCA TTTAGTATAA AATATTCAAG CCTAATTAGA AATATTCAAG CCTAATTAG AATATCAGG CCTAATTAG ATATTCAG CATTAGGAT TATAACGGAG ATGGTTCGCT AATATCAG CATTAGCAT TTCAGTATGAT AATATTCAAG CATTAGCAT TCAGACGCCC AATTAGCT AATATTCAGAA AATATTAGTT AATATTAGTAGAAA AATATTAGTT AATATTAGTA	TOTCANCTIT TCCAGGATA TCCAGGATA ACTGGCTT TCTGCAGCA ACTTTGAAA TGCAGGTAC GATANATGT GATGAGTAG ATTGGACAA ATTGGACAA ATTGGACAA ATTGGACAA ATTGGACAA ATTGGACAA ATGGACAA TTACGTAGG ATTAGGACAA TTACCTIGCA GATAATGGAT TACCTIGCA GATAATGGAT TTACCTIGCA GATTAATGAT TTACGAGAGAT TTACCTIGCA GATTAATGAA TTACAGAGAT TTACCTIGCA GATTAATGAA TTAGAAAA TTAGAAGAT TTAGGAGAGT TTTACCTTGCAGC GCAAACAGGT TTTAGGACAT TTTAGGAGT TTTAGGAGT ATTTAGGACAT TTTAGGAGT TTTAGGAGT ATTTAGGACA TTTAGGAGT TTTAGGAGT ACTTGACCAT TTTAGGAGT CACTGCACC CACTGCCC CACTGCCC CACTGCCC CACTGCCC TACACACACA TACCCCTTGGTCC TACACACACAA TTCCCCTTGGTCC TACACACACAC TTCCCCTTGGTCC TACACACACAC TTCCCCTTAGGACA TTCCCCTTAGGACA TTCCCCCTTAGGCAC TTCCCCCTTAGCAC TTCCCCCCCCC TTCCCCCCCCCC	ATCTCAGGTG TGTTGGAMA ATAAGTGGTG TGTGGGACCTGG TTGGGCTTCTA TTAAACAGT TTGTTCAGGG GCACTAGGTA ATAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT TTTTGTATGG AAAATGCAT AGCATGGAT AGCATGGAT AGCATGGAT AGCATGGAT ATCAAGTGC TTAGAATTAAGATTAC TAAGGAAGA AAAAGAAGTA TAAAGAATTA AGCATGGAT ATCAAGTGC TTAGAATTAAGATGC TTAGAATTAAGTTCA TAAGGAAGTA AGCATGGAT TTAAAGGTTC TTAGAAGTTC TTAGAAGTTC TTAGAAGTTC TTAGAAGTTC TTAGAAGTTC TTAGAAGTTC TTAGAAGTTC TTAGAAGTTC TTAGAAGTTC TAGAAGTTC TAGAAGTTC TAGAAGTTC TAGAAGTTC TAGAAGGAAATTA AGCTATTTAAG CCACGGTCT TCAGAGGCAA TTCATGAGT TCTTGAGATG TATTTCATTTAG CCTCTGGGC CAAGATCAT TTTTATAGTTTA CCTTTGAGATG TATTTATAGTTTC CATTTGAGATG TATTTATACATT CCTTTGAGATG TATTTAATAGTTTC CATTTGAGATG TATTTAATAGTTTAGATAGATT CCTTTGAGATG TATTTAATAGTTTAGATAGATT CATTTGAGATG TATTTAATAGTTTAGATAGATT CATTTCATCAGTT CATTTGAGATG TATTTAATAGTTTAGATAGATT CATTTCATCAGTT CATTTCATCAGT CATTCAGT CATTTCATCAGT CATTTCAGT CATTTCAGT CATTTCAGT CATTTCAGT CATTTCAGT CATTTCAGT CATTTCAGT CATTTCAGT CATTTCAGT CATTC	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAATGA AAAGGAGTTT TCTGAAAAGC CTTAAAATGT CTTCAAAAGC TTCAATAATGT TTCATATGAGGTT TTCATATGAGGTT TTCATATGAGGTAT TTCATATGAGGTAT TTCATATGAGGTAT TATGAGCACA AGTTCAAATGT TCAGAGCACAGT TCAGGAGCAGT TCAGGAGCAGT TCAGGAGCAGT TCAGGAGCAGT TCAGGAGCAGT TCAGGAGCAGT TCAGGAGCAGT TCAGGAGCAGT TCAGGAGCAGT TCAGGAGTAGT AACCAGCATAA AACCAGCATAA AACCAGCATAA AGGAGTTGTG TCAGGAGCTAGC AGGAAAATTTGC GAGAAAATTGG TCAGGAGCTAT GAGAAAATTGG TCAGGAGCTAC CTCTACTGAC CACAGAGCAAA CTCCAGCAGC CTCTACTGAC CACAGAGCAAA CTCAGAAAATTGC CTCAGGAGCTAC CTCAGGAGCTAC CTCAGGAGCTAC CTCAGGAGCTAC CTCTACTGACACAC CCTTTTTACAGAAAA TCCCAGAACATAC CCTTTTTACAGAAAAATTAC CCTTTTTACAGAAAAATTAC CCTTTTTACAGAAAAATTAC CCTTTTTACAGAAAAATTAC CCTTTTTACAGAAAAATTAC CCTTTTTACAGAAAAATTACAGAAAAA TCCCAAACCTG CTCAGAACAT TCCCAAACCTG CTCAGAACAT TCCCAAACCTG CTCAAAACCTG CTCAAAACCTG CTCAAAACCTG CTCAAAACCTG CTCAAAACCTG CCAAAACCTAC CTCAAAACCTAC CTCAAAACACAC CCAAAACCTAC CTCAAAACACAC CCAAAACCTAC CTCAAAACACAC CCAAAACCTAC CTCAAAACACAC CCAAAACCTAC CTCAAAACACAC CCAAAACCTAC CTCAAAACACAC CCAAAACACAC CCAAACACACAC CCAAAACACAC CCAAAACACAC CCAAAACACAC CCAAAACACAC CCAAAACACACAC CCAAAACACACAC CC	1111 1211 1411 1511 1611 1711 1911 2011 2011 2211 2311 2411 2511 2611 2911 3011 3011 3011 3011 3011 4011 4011 40
CHAMATOL  GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G COTTGEGA A COCTTCAGGA A ATGGGATA A ATGGGATA A ATGGGATA A ATGGGATA A ATGGARANG ACTAGARA G ACTAGAR	GANATANCGA GCANGGCCAT GCANGGCCATG ACCTGANGGG ACCGGGCCAGG FTGTANAGTG ACCAGATTTAT GGCATTTTAT GGCATTTATA GGCATTTAT ACGGGCCAGG ACAGAGGAGG ACAGAGGAGG ACAGAGGAG ACAGAGGAG CATAAGGAG CATAAGGAG CATAAGGAG CATAAGGAG CATAAGGAG CATAAGGAG CATAAGGAGAC TAAGAGGAG CATAAGAGAC TAAGAGGAG CATAAGAGGA CATAAGAGGA CATAAGAGGA CATAAGAGGA CATAAAGAGC TAAAAGAGC TAAAAGAGC CATAAAGAGC CATAAAGAGC CATAAAGAGC CATAAAGAGC CATAAAGAGC CATAAAGAGC CAGAGGAGAG GCCACAATA CAGGCAACA CTGAGGAGAGG CAGAGAGGG CAGAGAGG CAGAGAGGG CAGAGAGGG CAGAGAGGG CAGAGAGGG CAGAGAGGG CAGAGAGGG CAGAGAGGG CAGAGGG CAGAGAGGG CAGAGGG CAGAGGG CAGAGGG CAGAGGAGG CAGAGGAGAGG CAGAGGAGAGG CAGAGAGG CAGAGGAGAGG CAGAGGAGAGG CAGAGGAGAGG CAGAGAGG CAGAGGAGAGAG CAGAGGAGAG CAGAGAGG CAGAGAGG CAGAGAGG CAGAGGAGAG CAGAGGAGAG CAGAGGAGAG CAGAGGAGAG CAGAGAGG CAGAGAGG CAGAGAGAG	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATT TATGCATCA AATGCGGAAA CCCTTTGATC AATGGGGAAA GCAAAGAAGG TAACATGAT TAGGACTATACA TTTGGACT TAGGACTATACA TTTGGACT TAGGACTATACA TTTGGACTT TAGAACTC GCGGAACCAGA AATTATCA ACTGTGAA AATAAGAT TTCATATT ACACCTGTGA TTCATATT ACACCTGTGA TTCATATT ACACCTGTGA TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATT TGAGGACT TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATAT TGGAGCT TTCATATAT ACACTGGAT AGATGCT AGATGCT AGATGCT AGATGCT AGATGACT AGATGACT AGATGAATG ACTCATAAA AGGTAAAAGA ACTCATAAA AAGGAACAAC AAGGAACAAC AACCTTAATGG AACACTGTAG AATACACCCA AATACACTCA AATACACCCA AATACACTCAC AATACAC	TAGAGAGGTT TACAGAGGTT TCATTAGGTT CTTCAGATGC TACAGATGCCC TAAGGATACCCC TAAGGATACCCC TAAGGATACAC AGAGTGTCCTT TAAAGCAGCA ACATTGATAG ACCATCATCGTG GCTCATTTA ATTATCTCCTC GTCACTGCC ACACTGCCATTAAATAT TATACTTCCTTC TAAAGCATCACAA ATTATATATTCTCTCTT TAAGCTATAAATAT TATACTTCCTT TAAGATTAAATAT TATAGATCATACAAA TCATTAAATAT ACTCAGTAAAA TTAAGATTAATAT TAAGATTAATAT TAAGATTAATAT TAAGATTAAAATAT TAAGATTAAAATAT TAAGATTAAAATAT TAAGATTAAAATAT TAAGATTAAAATAT TAAGATTAACAAA ACTCAGGAAACCC GAACCTAAACAAA TCAGAGTAAA ACTCAGGAAACCC GAACCTGGGA AAAAGTCTA GGAAAACCCC GAACCTGGGA TCCAGGCTAAC TCCAGGCTAAC TCCAGGCTAAC TCCAGGCTAAC TCCAGGCTAAC TCCAGGCTAAC TCCAGGCTAAC TCCAGGCTAAC TATAGATTAC TCCAGGCTAAC TCCAGGCTAC TCCAGGCTAAC TCCAGCCTACAC TCCAGGCTAAC TCCAGGCTAAC TCCAGGCTAAC TCCAGGCTAAC TCCAGGCTACAC TCCAGGCTAC TCCAGGCC TCCAGGCTAC TCCAGGCC TCCAGGCTAC TCCAGGCC TCCAGGCTAC TCC	CAGAAAGCTE AAACTATGCAA GACATCTGGG TGGGGAAAGG TGGGGAAAGG ACATATAAT CAGTATTAAT TGCCTTTTAGG CATTCAGACA TGAGGCAGG AGAGAATAA CAAAGACTCA ACAGGCAGGA CAGAGACATAA ACAGTACTAGACA TATGGAGGA CTGGCTACAT TATTATAAA CCTGATTGTGG CACATTAA TATTATAAC TCAATTAAAC TCAATTAACT TACAGCATAAAT TATGTGGGG GGCTAGATTA TATTATAAC TCAATTAACT TACAGCATAAAT TATGTGGGG GGCTAGATTA CGCTATATAT CGCATTAAAT TATGTGGGG GGCTAGATAAAT CAAATGAGA CAATGAAAA CCAATAAAT CAATGAGA CAATGAAAA CCAATGAAAA CCAATGAAC TCTTCAGTC AACAATGGGG TACATACCTA CAGTGGTATTA	ACTGECCCTE AGTGGTAAA AGGACAAAGC ATAGATAAGA AAGGACAAAGC TTATTATTGC TTATTATTGC TTATTATTGC TTATTATTGC TTATTATTGC TTATTATTGC GCTTTGCAAG GCAGACTCT TCAGGGTT TCAGGGTAAAA TCAGGCAACA AGGCTAGAAC TACAGAGGT AATACAGAAGT AGGCTAGAT CAATATACAA CAAATACAA GCAGTTAAA AGGCATTATAA CAGAAAAAA CAGAAAAAAA CAGAAAGTTC CTGTTATTAT TGTGTTATTTATTATTGTGTTATTATTATTGTGTTATTAT	TGGTTCCCAC TGGTAGGGAT ATCCAAAGCCT ATAAGGATGA GCAATGTGAT ACAGTTTCGCTAT ACAGTTTCG ACAACTACGG CAATAATGTGAT ACAGTTTAGGAGA ACAGTAAGGAC CGCTAATGAA ACAGTAAGGAG TTTTAGTATG GATCAACAGAA AGGTTAGGAGA AGTTAGGAGA AGTTAGAGTA ACATAGAAG CCTAATTATA CATAGCATGAA ACATAGAAG ATTTTCATTAGAAG ATTTTCATTAGAAG ATTTTCATTAGAAG ATTTTCATTAAAA ACATAGAAG ATTTTGGTG AATATTCAGAG ATGATTGAT AATATTCAGAG ATGATTAGAAG ATGATTAGATA ACATAGAATA ACATAGAATAAAAA AGATTAGATAAAAAATGAAAAAATGATAAAAAATGAAAAAA	TOTCAACTIT TCCAGGCATC ANAGAAAATT TCTCTATAA ACCTGGCTTT TCTGCAGCCA AAATTTGAAA TGCCAGGTAC GATAAAATGT GATGAGTAGG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG CAGTTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA TTACCTTGCA GAGTAACAA GAGTAGATG CTGTGGTAAG TTGTCCCAGC CAGTAACAAA TAAATGTGAT AAAAAATGTGAT AAAAAAATGTGAT AAAAAAATGTGAT AAAAAAATGTGAT AAAAAAATGTGAT AAAAAAATGTGAT CCATGGCCAGC CATGGCCAGC CATGGCCCC CACTGCCCTAC AAAGGAACTT CCATTGCCCTAC AAAGGAACTT CCATTGCCCTAG AGGTACAGCA AGGTACAGTG CATTGCCCTAG AGGTACAGCA AGGTACAGTC TACACACACA AGGTACACCC AAGGTACAGTC TACACACACA AGGTACAGTC TACACACACA AGGTACAGTC TACACACACA AGGTACAGTC TCCCTTAG AGGTACAGTC TACACACACA AGGTACAGTC TTCCCTTAG TTCATCCTTAG AGGTACAGTC TTCATGTT TTCATCCTTAG AGGTACAGTC TTCATCCTTAG AGGTACAGTC TTCATCCTTAG TTCATCCTTAG AGGTACAGTC TTCATCTTAGTC TTCATCTTAG	ATCTCAGGTG TGTTGGANA ATANGTGGTG TGGGGCTTGTA TTANANCAGT TGTGTCAGG GCACTAGGTA TGTGTCAGG ANANTGCAT TGTGTCAGG ANANGAGGG ANANGAGGGG TTGATGTATAG ANGCCTTAG ANGCAGGATA TANAGAAGGAT TTANAGTAG TTANAGTAGAT TCACAGTCC TTANAGTAGAT TCACAGTCC TTCACAGTCC TTCACAGTCC TTCACAGTCC TTCACAGTCC TCACAGTCC TCACAGTCC TCACAGTCC TCACAGTCC TCACAGTCC TCACAGTCC TCACAGTCC TCACAGTCC TCACAGTCC TTCACAGTCC TTCACAGTCC TCACAGTCC TCACAGTC	TGAATCAGA CTCACAAAGC AGATGATAAA GAATAATGA AAAGGAGTTT TCTGTAAAAC TACTAGGGGT TCATAGGGGT TCCATATGG ATGAGGTAT TCCATATGG ATGAGGTAT TCCATATGG ATGAGGTAT TCCATATGA AGTTCAAATGT TATGGAGTAT TATGACACA AGTTCAAATGA TTATGACACA CCTCTATGAA ATGAGCACAGT TCAGGAGTAT AGAGCACAGT TTATGACACA ATTATATAAGG TCAGAGTAT AGAGCTAC AGAAACTAGA AGAAACTAGA AGAAACTAGA AGAAACTAGA AGAAACTAGA CCGCTCTCT AGAAAATATGA CCAGAGGCAAA AGAAACTAGA CCGCTCTCT TCAGGAGCAA AAACAGCTCC TCATACACAC CCTTTTACT TCAGGAGCAA CCTTTTACT TCAGGAGCAA TCCAAACACAC CCTTTTTACT TCCGAACAT TCCGAACAT ATTATAACCTT TCCGAACAT ATTATAACCTT TCCGAACAT TCCAAACCT TCGAACAT TCCAAACCT TCCAAACC TCCAACAT TCCAAACCT TCCAAACC TCCAACAT TCCAAACCT TCCAACAT TCCAACCT TCCAACCT TCCAACC TCCACAC TCCAACC	1111 1211 1411 1511 1611 1771 1811 1911 2011 2211 2211 2311 2411 2511 2611 2711 3011 3111 3211 3311 4011 4111 4211 4311 4411 4511 4611 4711 5011 5111 5211
CELAMATOR O GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G GOTTOTAMA G COTTGEGA A ACCOTTGEGA A ATAGAGAMA A ATAGAGAMA A ATAGAGAMA G ACTAGAGAMA A ATAGAGAMA G ACTAGAGAA G ACTAGAGAA G ACTAGAGAA G ACTAGAGAGA G ATATAAGGA ATATAAGGA ATATAAGGA ATATAAGGA ATATAAGGA ATATAAGGA ATATAAGGA ATATAAGGA ATATAAGGA ACAGAAGTA G TATACTTGACT TATACTTGACT TATACTTGACT TATACTTGACT TATACTTGACT TATACTTGACT TATACTTGACT TATACTGACT TATACTTGACT	GAMATAACGA GCAAGGCCAT GCATGAAGGC ACCTGAAGGG ACCGGGCCAGG ACCAGACTTA GGCATTTAT GGCATTTAT GGCATTTAT GGCATTTAT GGCATTTAT ACGGGATTTAT ACGGGATTTAT ACGGGAGAGGA ACAGAAGGAG ACAGAAGGAG ACAGAAGCAT ACGGGAGAGCAT ACGGGAGGAGAG ACAGAGGAGAGCAT ACGGGAGGAGAGCAT ACGGGAGGAGAGGA	TAMAMAMA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATT TATGCATC TATGCATCA CCCTTTGATC AATGGGGAAA GCAAAGAAGG TAACACTGAT TGGAGCTTTA TGGAGCTTTA AGTGAGCCT TAGGACTATACA TTTGGAACTC GCGGAACCAGA AATTATCCA GCGGAACCAGA AATTATTCA TTCATTTATATA ACACTTGGA TTCATGAT TTCATTATATA ACACTTGGA TTCATTATATA ACACTTGGA TAGATGCT TAGATGCT TTCATTATATT TGGAGCTGT TAGATGCT TAGATGCT TAGATGCT TAGATGCT AGAGCTGCAA AGGTGAAATG AGGTGAAATG AGGTGAAATG AGGTAAAAG ACCTAAAAA AAGGAACAC TATACAGCAA AACACTTGAA AACACTTGAAA AACACTTGAAA AACACTTGAAAAA AACACTTAAAAA AACACTTATAAA	TAGAGAGGTT TACAGAGGTT TCATTAGGTGT CTTCAGATGC TACAGATGCCC TAAGGATAGCA TAGAGATAGCA AGAGTGCTGT TAAAGCAGCA ACAATGATAG ACACTCGTG GCCTGATTA ATTATCTGCG GCTGATTA ATTATCTGCT TAACATCTGTG GCTCACTGTG GCTGATTA ATTATCTGCTA TACCTATTA ATTATCTTCTGT TACCTATTA ATTATCTTCTGT TACCTATTA ACTATCATCA TACACTATA ACTAGAGTAAA TTAAGATTAT TAGAGTAAA TTAAGATTAT ACTAGAGAAC ATAGAGTAAA ACTAGAGAAC ACACAGATAAA ACTAGAGAAC ACACAGATAAA ACTAGAGGAAC ACACAGATAAA ACTAGAGGAAC CCCTAAACAA ACTAGAGGAAC TGCATCATAA ACTAGAGAAC TGCATCATAA ACTAGAGAAC TGCATCATAA TCATTAGACAA TCATTAGACAA TCATTACACA TCAATCATTACACA TCAATCATTACACA TCAATCATTACACA TCAATTACACA TCAATTACACA TCAATTACACA TCAATTACACA TCAATTACACA TCAATTACACA TCAATTACACA TCAATTACACA TCAATTACACA TATATATGTG	CAGAAAGCTE AAACTAGCAA GACATCTGGG TGGGGAAAGG TGGGGAAAGG TGGGGAAAGA TGAGTATTAAT TGCCTTTTAGG CATTCAGACA TGAGGCAGG AAAGATCTCA GGAGTCCCAG GGAGACATAA TATGGAAGAG CAGAGTCTGA TTTTTGGAAGAG CAGAGTCTGA AAACAGTAGT TTTTTTGAC TTTTTTTGAC TTTTTTTGAC TTTTTTTGAC TGGACATTAA TATTATAAA CCTAATTATAC TACAGCATAA TATGTGGTG GACTTTAATTAC GCATTAAAT TATGTGGTG GGCTAAATT CAATTAAAC TCAATTAAAC TCAATTAAAC TCAATTAAAC TCAATTAAAC TCAATTAAAC TCAATTAAC GGCTTACATA GGCTTACATA GGCTTACAT GGCTTACT GGCTTACT GGCTTACT CACTGAGGT TACATCCTA CAATGAAAA CCAATGAAAA CCAATGAAAA CCAATGAAAA CCAATGAAAA CCAATGAAAA CCAATGAAAA CCAATGAACCC TTCTTGAGTC AACAATGCGTATT TACATTATCCTA TACATTATT	ACTGGCCCTG AGTGGTAAA AGGACAAAGC ATAGATAAGA AAGGACAAAGC TTATTATTGC TTACTGAGTT TCTGGAGTT TCTGGAGTT TCTGGCTACA AGGCTAGAAAG CAGAGTCATTA TCTGGCAACA TCTGGCAACA AGGCTAGAAC CAATAGCCA ATGCCTATTA ATGCCTACA AGGCTAGAAC TACAGCAGG ATGCCTTTTTAGA AGGCTAGAG AGGCTAGAG AGGCTAGAG AGGCTACAAA AGGCAAAAGCAC TACTGTAAA CCAATATCCT ACAATAGCA AAGTACTAAA AAGTCATATTA AAGCCATTTAAA AAGCCATTTAAA CCAATATCAG CCACCATTTAAA AGGCACAA TAAGGGAGGA AAGAAAAAA TAAGGGAAGA CAGAAAAAAA TAAGGGAAGA CAGAAAAAAA TAAGGGAAGA CAGAAAAAAA TAAGTATTATA CTGTTATTAA ATTATTCTGT TGTTTATTAA CTGTTTATTAA CTGTTTATTAA TTGTTTATTAT TTGTTTATTTA TTGTTTATTAT TTGTTTATTATTAT TTGTTTATTATATAT TTGTTTATTATATATA	TGGTTCCCAC TGGTAGGGAT ATCCAAACCC ATAAGGATGA GCAATGTGAT GCAATGTGAT ACAGTTTGG ACAACTACGG CAATAATGT ACCATTAGGAT ACCAGTTAGGA ACACTAGGG CCTAATGAA ACCGTAGGT ACCTAAGGAG ACTTAGGAG AGTTAGGAG AGTTAGGAG AGTTAGGAG AGTTAGGAG ATATTAATT AACACTAGAG CCTAATTAAA ACATAGAAG CCTAATTAATT AATATTCAG CATAGCATT TGATTGAT CATTTAGTAG CATAGCATT TAATAGAAG CATAGCATT TAATACAGAG ATGGATGCTA AATATTAGTG AGGATGCTA AATATTAGTA AATATTAGAA ACATAGAAT AACTTAATAAA AACATAGAAA TGATTATTAAAA AACATAGAAA TGATTATTAAAA AACATAGAAA TGATTATTACA AGAAATGTGT ATGGCCCTA ATGTTAATAAAA AACATAGAAA TGATTATTACA AACATAGAAA TGATTATTCTC TACACACACA	TOTCANCTIT TCCAGGCATG ANAGANANTI TTCTTCTATA ACCTGGCTTT TCTGCAGCCA ANATTTGANA TGGCAGGTAC GATANANTGT GATGAGTAGG ATTTCAGGAG ATTTCAGGAG ATTTCAGGAG ATTGAGTAGG ATTGAGTAGG ATTGAGTAGG GAGTACAAA GACTTTGTGT TTACCTTGGC GAGTAGAGTA	ATGTCAGGTG TGTTGGANA ATANGTGGTG TGGGGCTTGTA TTANANCAGT TGTGTCAGG CCACTAGGTA AAATGCAT TGTGTCAGG AAAATGCAT TGTGTCAGG AAAATGCAT TGTGTCAGG AAAATGCAT TGTGTCAGG AAAATGCAT AGGAGGGGC AAAAGAGGGG CCAAGTTAC AATCTGAAAT AGCATTGAAT AGCATGAAT ATTAAGTAC TTAAGGAGGAT TTAAGGATGA ATTAAGGTTC TTAGAGTTC TTCAAGTC TCCAGGGCCA ATATTGAAT TTCTTCAT TCTTCAGGTC CAGCCTGGC CAGCGCTGC CAGCAGGAAG TCTTCAGGTC TCTTCAGGTC TCTTCAGGTC TCTTCAGGTC TCTTCAGGTC TTAAGTTCATT TCTTCATGTT TCTTCAGGTC AAAATACAGTA TTTTATTGATA TTTCATCATT TCTTCGGGC AAAATACAGTA ATATTGATAT	TGAATCAGA CTCACAAAGC AGATGATGAA GAATAATGA AAAGGAGTTT TCTGTAAAAC TACTAAGGGGT CTAAAGGAGTTT TCCATATGG TCCATATGG ATGCAGATA TGCATATGG ATGCAGAGA AGTTCAAATGT TACGAGAGACAC CCTCTAATGG TATGGAGTAC CCTCTATGAA ATGACACAC TCAGGAGACAC ATGAACACAC ATGAACACAC AGAGCTATA AGAGCTAC AGAACACT CAGAGCACAC TCAGGAGCAC AGAACACAC CCTCTACTGAA AGAGCTTCT CAGAACACAC AGAACACAC CCTCTACTGAA AGAACTCAG CCAGACCAC AGAAACTACAC CCAGACCAC CCATACACAC CCAGACCAC CCAGACCAC CCATACACAC CCAGACCAC CCAGACCAC CCAGACCAC CCATACACAC CCAGACCAC CCATACACAC CCAGACCAC CCATACACAC CCAGACCAC CCATACACAC CCAGACCAC CCATACACAC CCAGACCAC CCAGACCAC CCATACACAC CCAGACCAC CCACAC	1111 1211 1211 1311 1411 1611 1711 1911 2011 2011 2011 2011 2011 2011 20
CHAMATOR OF CAGACATA A GOTTOTAGA A CONTIGORA A ATAGAGATA A ATAGAGAGA A ATATAGAGA A ATATAGAGA A ATATAGAGA A ATATAGAGA A AGATAATAGA A AGATAATA C CAGTAATA C CAGTAGATA C CAGTAGATA C CAGTAGATA C CAGTAGATA C CAGTAGAGA A GOTTACAGA C CAGTAGATA C CAGTAGAGA C CAGTAGATA C CAGTAGATA C CAGTAGATA C CAGTAGATA C CAGTAGATA C CAGTAGAGA C CAGTAGATA C CAGTAGAGAC C C C CAGTAGAGAC C C C C C C C C C C C C C C C C C	GAAATAACGA GCAAGGCCATG GCATGAAGGC ACCTGAAGGG ACGGGCCAGG ACGAGCCATTA GGGATTTAT GGCATTTATAA GTAGAGGA ACTAGAGAGT CTATATATAA GCAGACTCT ACTAGAAGT ACTAGAGAGT CTATATATAA GCAGAAGCAT ACTAGAAGT GCAGAAGCAT ACTAGAAGCAG ACTAGAGAGC ATAAACTGA GCAGAAGCAT ACTAGAAGCAT ACTAGAAGCAT ACTAGAAGCAT ACTAGAAGCAT ACTAGAAGCAT ACTAGAAACG ACTACCAGA ACGACCAATT ACGGCCACAATT ACGGCCACAATT ACGGCCACAATT ACGGCCACAATT ACGGCCACAATT ACGGCCACAATT ACGGCCACAATT ACGGCCACAATT ACGGCCACAATT ACGGCCACCAATT ACGGCCACCACCACCCACCCCAC	TAMAMAAA TAGATATATC CCCTTGAAGA AAGTTAAGTA GGAATTTTC TGTTTGCATT TATGCATCA AATGCGGAAA CCCTTTGATC AATGGGGAAA GCAAAGAAGG TAACATGAT TAGGACTATACA TTTGGACT TAGGACTATACA TTTGGACT TAGGACTATACA TTTGGACTT TAGAACTC GCGGAACCAGA AATTATCA ACTGTGAA AATAAGAT TTCATATT ACACCTGTGA TTCATATT ACACCTGTGA TTCATATT ACACCTGTGA TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATT TGAGGACT TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATT TGGAGCT TTCATATAT TGGAGCT TTCATATAT ACACTGGAT AGATGCT AGATGCT AGATGCT AGATGCT AGATGACT AGATGACT AGATGAATG ACTCATAAA AGGTAAAAGA ACTCATAAA AAGGAACAAC AAGGAACAAC AACCTTAATGG AACACTGTAG AATACACCCA AATACACTCA AATACACCCA AATACACTCAC AATACAC	TAGAGAGGTT TACAGAGGTT TCATAAGGTT CTTCAGATGC TACAGATGC TACAGATCCCT TAGAGATACAC AGATGTCCTT TAAAGCAGCAC AGATGTCCTT TAAAGCAGCAC ACAGTGCTATAACAACAACAACAACAACAACAACAACATGCCTAACAACAACACACAC	CAGAAGCTA GACATCTAGGA GACATCTAGGA GACATCTAGGA GACATCTAGGA TGGGGAAAGA AGATAATACT CAGTATTAAT TGCCTTTAGG CATTCAGACA TCAGGCCAGG GAGACATAA ATCAGATAAT ATATGCCCAG GGAGACATAA TCTGTGGGA TCTGTGGGA TCTGTGGGA TCTGTGGGA TCTGTTGGA TATTTATAAAA CCTGATTTTA TATTATAAAA TCACATAAT TATCTGTGGGG CACATTAACT TACAGCTAAAT TATCTGTGGGG CACATTAACT TACAGCTAAAT TATCTGTGGGG CACATTAACT TACAGCTATAAT TATCTGTGGGG CACATAAAT CCAAATACT GGCTTAAAT TATCTGTGGG TACAATACT CACAATACT CACAATACT ACAATCGGG TACAATCCGG TACAATCCGG TACAATCCGG TACAATCCGG TACAATCCAT TACATTATT TACATATTATA TATATATA	ACTGGCCCTG AGTGGTAAA AGGACAAAGC ATAGATAAGA AAGGAGAAAGC TTAATTATTGC TTACTGAGTT TTACTGAGTT TTACTGAGTT TTACTGAGTT TTACTGAGTT TTACTGAGTT TTACTGAGTG GATTTGCAAG GAGTCATTA ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA ATCCTGAATA AGGCCATCGA ATGCCTTGTT AGGGTTACAG AGGATTTCCT ACAATAGCA TACTGTTAAA CTATTTCTGT GCAGTCAAGC TACTGTTAAA CTATTTCTGT ACAATAGCA TACTGTTAAA CAATATTCTG TGCAGCACT TACTGTTAAA CAATATCAAA AAGCCATTTT GGATCAGGC CCCAGCACTT AAAATACAAA TAAATACAAA TAAGGGAAAAAA TAAGGGAAAAAA TAAGTGAGC CCTGTTATTA ATTATTCTG GGTCGCCCAAC TACTGTTATTA ATTATTCTG GGTCGACCAAA TATACACAC TTGTGTATTCTA TTGTGCACCACA TATACACACA TATACACACA TATACACACA TATACACACA TATACACACA TATACACACA TATACACACA TATACACACA TATACACACA	TGGTTCCCAC TGGTAGGGAT ATCCAAAGCC ATAAGGATGA GCAATGTGAT GCAATGTGAT ACAGTTTAGGATGA ACAGTTAGGAT ACAGTTAGGAT ACAGTTAGGAT ACAGTTAGGAT ACCTAAGGAG CGCTAATGAA ACAGTAGAG TTTTAGTATG GATCAAGTAG GGTTCGGGAT TGGTTCAGCAT AGGTTCAGCAT AGGTTCAGCAT AGGTTCAGTA ACATTAGAAG CCTAATTAAC CATAGCATTAT TATATCAGAG AATATTCAAG CATATGCAT TATATCAGAG AATATTCAAG CATATGCAT TATATCAGAG AATATTCAAG CATATGCAT TATATCAGAG AATATTCAAG CATATGCAT TATATCAGAG AATATTACAAG CATATGCAT TATATCAGAG AATATTACAAG CATATGCAT TATATCAGAG AATATTACAAAA AACATTAGAT TATATCAGAG AATATTACAAAA AGAAATCTGT ATCTTATAAAAA AACATTAAAAA TGATTATCTT TATATCAGAAA TGATTATCTT TATATCAGAAAA TGATTATCTT TATATCAGAAAA TGATTATCTT TATATCACACA TATATCTATA	TOTCANCTIT TCCAGACATE TACAGGAAAATT TTTCTTCATAT ACCTGGAGCCA TAAACTCTCA AAATTTGAAA TGCCAGGTAC GATAAAATGT GATGAGTAG ATTTGAGAG ATTTCAGGAG ATTGAGAAATGT TACCTTGCA AGATACACAA AGATCACAA AGATCACAA AGATCAAAAGTT TACCTGGAAAGTT TACCTGGAAAGTT TACCTGGAAAGTT TACCTAGGAAAGTT TACCTAGGAAAGTT TACCTAGGAAAGTT TACTAGGAAAGTT TACTAGGAAAGTT TACTAGGAAAGTT TACTAGGAAAGTT TACTAGGAAAGTT TACTAGGAAAGTT TACTAGGAAA TAAAAGTGAT TACTAGGAAA TTAAAGTGAT TACTAGGAAA TTAAAGTGAT TACTAGGAG CAATGGGAG CAATGGAGT CACTCGGTT CACCCAGT CACCCAGT TACACACAC CATTAAAGTT TACTACACAC CATTAAAGTT TACTACACAC TACACACAC ATTAAAGTT TATATATATT TATATATGTA TATATATGTA TATATATGTA TATATATGTA	ATCTCAGGTG TGTTGGAAA ATAAGTGGTG TGGGGCCTGGG TTGGGCTTGTA TTGTTCAGGG CCACTAGGTA AGAGGGGGCC AGAGGGGGCC AGAGGGGGCC AGAGGGGGCC AGAGGGGGCC AGAGGGGGCC AGAGGGGGCC CCTAGGAAGTTA AGCATGGAAT AGCATGGAAT TTCAGGAAGT TTAGGAAGT AGCATGGAA ATCAAGTGC CCTAGCAAGT ATCAAGTGC TTAGAATTA AGGAAGGA ATTAAGGAAGA TCACAGGCCA TCCACAGGCCA TCCACAGGCCA TCCACAGGCCA TCCACAGGCCA TCCACAGGCCA TCCACAGGCCA TCCACAGGCCA TCCACAGGCCA TCTACAGGCCA TCTACAGGCCA TCTACAGGCCA TCTACAGGCCA TCTACAGGCCA TCCACAGGCCA TCCACAGGCCA TCTACAGGCCA TCTACAGGCCA TCTACAGGCCA TCCACAGGCCA TCTACAGGCCA TCTTCACAGGCCA TCTTCACAGGCCACAGA TCTTCACAGGCCA TCTTCACAGGCCA TCTTCACAGGCCACAGA TCTTCACAGGCCACACACACACACACACACACACACACAC	TGAATCAGA CTCACAAAGC AGATGATGAA GAAATAAGG AAAGGAGTTT TCTGTAAAAC TACTAGAGGT TCAATAGGGT TCAATATGG ATGGAGGTA TGGAGCAAGA AGTTAATAGG TATGGAGGTA TGTGAACAC CCTCTATGAA TGTGGACACA AGTTCAATAGG TCAGGAGCACAT TCAGGAGCAGA ATATATAGG TGAGACACAT ATATATAGG TGAGACACAT AGAGCTTCT TCAGGAGCTGC TCAGGAGCTGC TCAGGAGCTGC ACCAGCATAA ATATATAGG TGATATTTCC AGACACACAC CCTCTATGAA TTAGAGCAGC TCAGGCATTAA ATATATAGG TGATATTTCC GAGAAACCTG GAGAAACCTGG TGATATTTCC TCAGGACAC ACCAGCCT TCAGGAGCT TCAGATATATAT TCATATATAT TCATATATAT TCATATATAT	1111 1211 1411 1511 1611 1711 1911 2011 2011 2211 2211 2211 2311 2411 2511 2911 3011 3111 3211 3311 3411 3511 3411 3511 4411 4511 4611 4611 4611 5111 5111 5

### Figure 8B

AGATGTAGCT TAAAAAAAC ATATCCTGGA ATTCTAGAGA GATGCTTAAA TCACTGCAAT TCCTATAACA CTTGCCAACC AAAGGTGCTG TTGATCTGAA ATTGCTTTTT TAAATTAATG CAGTGATTTT TCTTTAACAT CTAGTGACAG ACACTGGGGT CACATTTGCA GCTGGACCAT AATTAGGCTT CTGTTCTTCC GTGATGACAA GTTTAAAGT ATTTGGGCAA CCATATTCTG AAAACAGCC AGCCAGGGTG ATGATCACT TTGCAAAGAT CCTCAATGAG CTATTTTCAA GTGATGACAA AGTGTGAAGT TAACCGCTCA ATTTGACTTC TTTCTTTTTC ATCCAAAGTA CTGTGGGAAC ATCACAGAT TTGGCCTCAT TATTACTGG ATTCTCTTGA CTAAAAGTAA AATTGAATT TAATTCCTAA ATCTCCATGT GTATACAGTA CTGTGGGAAC ATCACAGAT TTGGCCTCCAT GCCCTAAAGG GAAATTGGCT TTCAGATTAT TTGGATTAAA AACAAAGACT TTCTTAAGAG ATGTAAAATT TTCATGATGT TTTCTTTTTT GCTAAAACTA AAGAATTAT	5911 6011 6111
Val Phe Leu Asp Bis Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg Tyr Asn Ser Gly Lys Leu CTITTACATT TCAG IT ITT CTT GAT CAT GAA AAC GCC AAA ATT CTG AAT CGG CCA AAG AGG TAT AAT TCA GGT AAA TTG	6393
GAA GAG THE GIL GAA GGG AAC CTT GAG AGA GAA TGT ATG GAA GAA AAG TGT AGT TTT GAA GAA	
MAG ACT GAN AGA AGA G TGAGTATTIC CACATANTAC CCTTCAGATG CAGAGCATAG AATAGAAAAT CTTTAAAAAAG ACACTTCTCT TTAAAATTT  AAAGCATCCA TATATATTA TGTATGTTAA ATGTTATAAA AGATAGGAAA TCAATACCAA AACACTTTAG ATATTACCGT TAATTTGTCT TCTTTAATT	
The Glu Phe Tep Lys Clu Tyr Val TITATAG ACT GAA TIT TGG AAG CAG TAT GIT G G TAAGCAATTC ATTITATCCT CTAGCTAATA TATGAAACAT ATGAGAATTA TGTGGGTTT	r 6763
TICTCTGCAT AAATAGATAA TATATTAAAC ITTGTCAAAA GGACTCAGAA AGATCAGTCC AACCCTCTAA CCCATATTGG ATGGTGATAT ACTACAGGG TATGCCAGTG TGGGAACTAT CGCTGGTAAA TAAGTTTAAT CCTCCCTAGG GCTTCACAAA GAACATTGTT CCACCCCAGG AGGGTGGAAG GAAGAAACT AAATGATTGT GTCTTAGAAA CTAATGAAAG TTTGCATTCC TCAGTAAAAT CAGAGACTGC TGATTGACTT AAATGTTTAT AGCTTCAAAG TATCATGGCC CAGAAGCCCT TCCATGATTG TCCTTCCCCA CCCTCCCCAT TACCCTTCTT GCCTCCTCG CTACTTCTT CCTCGCACAC TGGGGTCCA	G 6963
CRECURAL CITECTURA CITETRAGA ALTERAGAA TECTECCACT TIGGACGOTT TATCTGGGGT TITCTCTTAT ITGGCTGTTC. CCAACTICG GTGGGCTGAC TCCCTCACCT CCTTCGGGTC TITGCCCCAAA TGTTACCATC TTAATGAGGC CTACCTTCAC CATCTATTAA TACTTCAACC TGCCCCAGT GCCTTACCAC TCTAGACACC TGTACAGAAC TCCACTCTAC TITTTAACA AGCTTTTCAC CATCTAATGT ATCATATAA TTCTTCATCA TACTATTAA	7263 Å- 7363
CATTATTIT CTCCTACTCC ACTAMATIC ARGITTCATG TIGGCAGGGA TATTCAATIG TITTGTITAT TGATATATIC CTAGCACCTA GAACAGTAT TIGGAAAAGA GATACTCAGTA GATATATTAT AAATATATA AAATATATA CAAATAAGAATA CAAAAAGAAAA AAAAAGAAAT TAAAAGAAAA TAAAAGAAAT TAAAAGAAAA TAAAAGAAAAT CAAAAAAGAAAAT TAAAAGAAAAT TAAAAGAAAAT CAAAAAGAAAAT CACAATACCAAAGACTAC CCAAACTCCA TCAAATGAAAA AAAACAAAAT CAAAACTCACCC CCAACTCCCA TCAAATGAAAA AAAACAAAAT AAAACAAAAT CAAAACTCACCC CCAACTCCCA TCAAATGAAAA AAAACAAAATCAA AACTCCTCCC	C 7563 A 7663
CTITIOTITIT TARATERIA GAGAMARGA GITGACTCTG TTATATTGTT TTATCTACCT TECCTTGATC TTAGARACGA ATACTACCAT ACCAGCTTC ACTGAGGTGC CCCCTARAGT INGTCCARAT AGGTCTTTGC AATCTCCATT CCCGCAGAT TTAGARCTTT GAATCACATG ATTTATTTCT ARANGTRAG CCATGCCGAT TTTCCCCACC ARAAAATTCC TGACTATAA ACTCCCTACAA TCCCTTCATT GCTFACTCCC CACCCCACG ATCATATTTT BACTTCCCC	T 7963 T 8063
CCTTGCCTTT TGGGTCACAT AGGTACACTG TTTGCTATAC CACAGGTATA GCTATCTGGA AAACATGGAG GGTATTATTC TGTTACTACT GCTTCGTCA CCAALAATA AAACAAAAAAAAAAAAAAAAAAAAAA	T 8363 C 8463
CAGGAGAGTA AACTGAAGCT TAGGGAAGTT AAAAGAACTG CCAAAAGCTCT CCCAGTTGGG GAGTCATGAA GCCCAGAAGA GAAGCCAAAT TCTCTGCTG TCAACCCCTT GCTTTCACTA TTACACCTCA GGGCCTTCAA TCTCATAATAGC AGTTATTCAT TAAACAGGAA CCTGGATAGTC TTAAACAGGA ATCTCTCAG TGGTAAAGATC TTGTCTCTTGT TTGTATTTGA CCCCAACTGT CTATGGCTTT GCCTGAACACA CAGCCAGAA ACCAAAAGGAG AACCAAATG GGGATAAAAT GACACTCATT TTAACGACAT GTCTCAGCAA ATGAGTTCCT GTGTAGCTGG CTGAAAAGCCC AGACCCTTTC AGTAAAACAT CCTGAATAA	C 8663 T 8763
TEACHTITAT TGGTUTATA TATANAGGG AAATGTAGCT CATTITTAGA CCAGITCTGA ACATCANTAG TAACAACCA GAGATAACCA ATTITGTTI CATAGAATTG GAACAAATTA GAGTATCTGT GCAAAAGCAT ATCAGATCTA GGAGCAGGG GGACAAGGT TAATTITTAA ATAAGCAAAT TITCCAGAG GGGACTACTT ATGATAAAGG GATATTAGTC TCTTAGTCAA CGGAACCTGG ATACACCGCT CTCACAGAGA AGAGGGAGAA TAGCCAGGAA TCTACAGAGA	9063 · A 9163 C 9263
AGATGTCAAG GAGATTTGCT TTAAAATACG ACTGATAATT AGAAATTTCT CAGTTTCCCC CTTTTCCCTC ATTCTTTGAT TCTTATTGTT ATCTTTATC CTTACTCCTT TGTTTCTCAT ATATTGAGTC TTACAGATCA AGGTCCCATT TTTTTCTTCA GGGGTATTTT TCTACTTCAA AGTGCCTACC ATCTCCCTT TGGTTCTATT CATCCTTCTC TCCCAAAGCT CCTTTACAAG TGTGGATTAA GGCAGAGCAC TAAGAAACCA GACTTAAAGA TTCCCTTCTC ATTCTGACT TTCTCCTTTC ACCTATTCCT TCCTCCTGTT TTCTTACCAT CAGTGTCTTC AAAGGCTTC AAGTACACGG TAAAATGCAGA AACTTCAAGA AAGGCAGAGCACAT	C 9463 T 9563 T 9663
GGAAACATAA CCAATGCATA CATAAATAAA GCACACTGTA GAATCTTTTT AAATTCTGTA TGATATATCG AATGCTGTCT CTCACATTAC CTAGACCAT TGAAACCGAA TITGTAAAAC AATGACTATC TITAAGTAGT AACAGATGCT TCTGACATGT TITCTATTGT CTTGAACAT TACTGCATAT GATACATCA AGTTAAGTGA CAATACAAGA AAGCAGATTC ATTTGCTCCC TGCCTAGGCC GTCAGTTCCT AAAGTGGAA CGCCATATAT TATCTAGCTC AGTTTGCTC ACAAGACCTG CAATAGAGCC TTGTGTGACA TAGAGATAAT ATTTGTTGAA GCCATTAAAT TTGACTTGGA ATTAACTCTG CCATCATTCT ATAAGGAAC	A 9863 T 9963 G 10063
ATTGAMANT CITCTCACC TOTGCTGATA TAGTACCTTT CTATACAAAA ACGTCCTTCT CCCTCTTCCC TTGGATTGCA TAAACTATGT ACATGCCTT CTCAGGGGGCA CTTTTCTAGG ACAGTGTCAG CCTAAGGATC TTTGTTTGGG TGGCTTTTAG AAACTCAGGA AGACAGGAGC ATCATATGCC TATAGGCAC TGGCTTCCAG GTCAGTAGTT TTGCTCTGAC CCTAAAATCA GACTCCCATC CCAATGAGTA TCTACAGGGG AGGACCGGGC ATTCTAAGCA GTTTACGTC	c 10261
Asp Gly Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp Asp II  CANTICANT TCTTAACCTA TCTCAAAG AT GGA GAT CAG TCT GAG TCC AAT CCA TCT TTA AAT GGC GGC AGT TCC AAG GAT GAC AT  Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys Asn Cys Glu Leu	10450
ANT TOO TAT GAN TOT TOG TOT COO TIT GGN TIT GAN GGN NAG AND TOT GAN TIN G OT ANGINACIAT TITTIGANIA CITCATGOT.  AMAGITICO: TOTGANACAN GITGANACTG GANANTGCAN TATTGGTGTA TOATANITIT TOTTANAMAC ATACOTTEGN TGCTTATANA CATTTCAT	T 10637
GTAGTGATAG TITTCAGGAT ATGAGTICAA GAAGCTACAT TAAAATCAAT AACAATATIT GGTAACTAAT ATTAAGTAAT AATGATGTIC CGACTCAC TATTAATCTI TAATACAACC GTATGTGGTT AGTACTATCA TTATGCGCAI TCTATGCAGA TAGAAAAACC GCAACTCCAA CGGCCAAAAA TTACGAGAG ATAAATGGTT TAGACAGGAC TITAAACTCA GTGTGACCAA AACCCATGCT TCTAACTACT ATATTCAAAA CTCAGAGAAA ACTGAACCCA GAAAATTG ATCATGACTA AATTGCTATC AACATAGGTG AAAGTCAATT AAGTACAGAA CTGGAGTATG ACTGGCCAAT TATCCCATAT AATGGGAATT CTCCACAT	GC 10837 NA 10937 GT 11037
ACAAACCACT TCATATGCTA AACTTGTTGA CAACATTCAA AGCTCATCCC TGAATTTGAC TATATTGATT ACATCGAAAA TGTTACATAG CAACCTTA ATCCTTGTTA ACCTTTTCTT CTCAAAGCCT AGATTATTTC TTTTTCCGAC GTTTTCAGTA ATTGGAGCAG TAAACCCCAG TGTCCCTTAC CTACTTGT ATTACCTCCA GATGCAATAT TACTGGTACT GTCATTGAGA AACGCACAA GTCCTAATGA GGAATTACACT TTCTACTCTG ACACTCTGGA AGAATAGA TGCAATCCTA AGGAAGAATT TAACACCACA GGCTACATGA CTAAGGATAA AGAGTAGAAA ATTAGCAGGA CTCTATTAAC CGATTACAGC AATCCACC	TT 11237 GA 11337 TG 11437
ACAGATGANA AAGGCATGAA ATGAAATGAA ATGTAGCAGC TACACTCGTC CTATTGAGAA AGGAAAAAAG TCACCTGTAA TGTTGTTCAG AAATCCTT AGTACTANA AATTCATTGA CCATCTTCCT TTAGTCTCGA AAATTTCTTA GAAGGTAAAA AAAGGAAAAAG GTGACAGGGC AAAGACATTT GAAAAGAA AAAAGAGTGA ATGAACTTGC ACACCTGGCT TGGACTCCCC ATTCCCCTTA GGTTTCCATT GTGGGGGACA AACTAATGCC TGGGTTACCT TTCTTGAG TGTGTTAATT GATTCAATAT CTCTGAAGTG CTACTTTCAT CTGAAAGGTT ATAATTTGAA ATTCAGATTT ACCTGGATAA ATTTGATCTT GCTATTAT	TC 11537 AG 11637 AG 11737
ANACCTCIAG MANTECTIGG AGTAGITACT CATTATCAGC TIMANTANIA TAGCCGGTGG AGCTGAGGGA ATGAGTAACT CAATTAGICT CAGITACA TGAAGGGCAC ATIGITOTAA ACTATAATIG AAAACATAAA TATCTITIACC TÄGITITAAA AATAAAGATG CITITAAAAGG AGGAAGGGAA TAGCCCCTG GAATGIAAAT ATAAGCACAA AACTICTACA ACAGAGTITG CIACGTGTGT GGCTGTGTTC CACCCAGGCAA AAATGCTAAG TCTACAACTG ACACAACT GAIACTCTCA TGTTCCCACA TITTGGTTTG GTCAAGGCTG TGCACTTGTA CTGCACGCCA CCACCACTCC TGGCCCTCTAC AGTATATTGA TCTGACCG	AC 11937 AG 12037 TG 12137
CANTETGATE AAGGTTTAGA AAATATTT CAGECEAGTT AGETEACAMA CAAAATGAGA ATTECCACAA ATTGCTCTTT ATCTEAGACA ACAGAGGA GETACAGCAA AAAGCATAAA TAAAGACAAA TAAAGACAGA ACAGATTAAT GGGGCATE GETACAGCAA AAATTACCAT TTAAGGTTTT TGCTACAAAAAAAAAA	AA 12337 TC 12437 TT 12537
ATGTCCCTAG CTGTACATCA AACCCCAAATA TCTCTCAGAT AAATGAAGGT CTGTAAGAAT TTGGTCATTC CTGTCTCTC TAAAGAGTAA CAGAGGCA TTCCCGGCAGT AAAGTAGAAT GGAAAGAAAA CAAAAATCAC AAGCCTATAA ACAACCTTCTT CAATTTTCCC AGCATGTCAC AGACACTACT GTCTTATT CTACGTATTT CTGAGGAGTA AAAAAAGGAA ATATCTTGAG	TT 12737 TA 12837
OCTITIGATIG TCACACAGT GGATATATCA GGAAATATAA AGGCAGAATA AACTAAAGCA CAACATACTA ACATITIGAG TAGGCATGAA GGGAATTA AAGTGTITGT GTTAACATGG AGGCAGGAG AACAGATGCT TIGAGATGT CITCAACAGA TATTCTAGGC ACTGAGACCC CCTTCGGGAC CAGAGAGA CCATATCCAC CACAGTACCT GACACATAAA TGGTCAGTAA TIGATAAATG AGTCCCATTC TAACTGTTCC TTAGCCCTGC TCTATGGAAC TCTCCCCT ATTCCTTCTG CCATTATTIT ATTCTGGAA TCTTCAGCCT TITAGCTCAG GGCAAAAGAT TGCTGATTAG GAAGCAATAT TTCCCACCTC CTGGCGAA CAAGCCAAAG ATCAACAGCA GCAGGAACAT ACTGAGCCCT AAAGGGCAAT GACAAATGTG GAGAATGATA CAGAGGTCTG GTTACTTCTT AGCCAATG	GA 1323; GA 1323; GA 1323; GA 1323;
ACAGANICAC ANTIGAGANA ACACAGAGIT TATTCATTCC CATTGTGCAT GCCCTGGACA AACCAAGCTG CACCTTTCGT AACTTATCAC AATCTCAT	AT 1353

# Figure 8C

TEACGGAACA CTTTCTACAG OTAATGTTTG ATTTGGCTGA ACACTITAGC ATTGCTTCGT AGCAACAAAA TGATAGCTAG TAACAGAAAA AGATCCAG ATATTACCAC TGTTAGTGAG GAGAAAGGCC TTTTAATTAA TTAATTAATT AATTAATAGG ACCAAGTGCC ATCTTTTTGG ATCATGCCCT TAGTGGAT	X 1363
TIGGTAGCAA AGGTTAAAGC TCAAGGTGGT TCCTTTGTCC CCCTGGCAAC AGTTGATTTG CCTCCCTTAT CTCCTGAAGT ACCGTAAGGA CTAAGAGC	
ATTATTALAT ITUGETATUS TAGSATATGI AAAATASASA TTAAAASASSA SISTEMA ASAASAA SAASAA TAGSATATGI AAAATASA AAAAA AAAAA	
CITAGILLAL LIVAGILAGUA AGAAGUGITG AAAGAATTING WYWCHCACTA WCWCHCWCC AFRICANTAGA ACTINGGACHE ARCACCAAN AAAAAAA	
ATGARGETIG TOCATAGGAG ETATOTALTA CAGTCACTCA TETETCALARTE TERCONTRA TOLUMCEACTA CACCACTCACTCACTCACTCACTCACTCACTCACT	
TETTANTUA GCAGGIUTAA GCTAACAAGT CETGAAACAT GCTACTTTTT GTTATTGGTA TTGCTATAGGA GAAACAAAGC GAAAGAACAA AAAmaa ma	
ATACANACAA GATGGCAGGA ATAGCCANA ANTATCAGGA ANCACANTTA TIGTGANTTG GGATTANACT ANTCTATTAN TANTGACANC TITCAGCT	
GAGTTANAN TITANTIGTA TACTGTTANC GAAGTGATA CCTANANTAN ANTIACACTG GCAGGCCANA ATGAAGGGAT GTGAAAAGAN CTATCAGG , MANCTANCA RANAGANACT AGCANAGCAN TCTTANTATC AGACANANTA GANTCCANGN GGNANTCAT TTCANANGAC AAGAGATTIT TITTATTA	FA 1493
ATTICATION UNITATION ATATOTAACO ACCTOAGOCA GOTTOAAGTO GOTTOTTOGO COTAATGOO TROCTAGOO AAGAAGTO	===,
ATCITITATE TIAGGITIAG GGGTALATUT GAAGGTTINGT TACATAGATA ALCANCINCT ACAGGGGTTT CTNCTLCAPA TRACADA AAAAA AAAA	
TENGETURGE AUGMANTAGE GATUITITES GETECTURG CTCATECAL CETECTURE CALCARDATE CALCARDATE CALCARDATE CONTRACTOR CONTRACTOR	
ATABUTUIT ARCAUTTAGE TUUGUTTAG ARGIGAGAE ETGCAGTATT TGATTTTCT TCCTAGGETA CTTVCCTAAG GARGAGACACACACACACACACACACACACACACACACAC	
TRAINING CACAAAGACAC ATAATCTCCT TCTTTTCTAT GGCTGCATAA TATTCCATGG TATATTATGAA CCACATTTTCTC TATATCCACA AACAAAAA	
TOBULATITA GGTTGATTCC ATGTCTGCTA TTCTAACACT GTAATTTCTA AAGACTTCCA GATTCTACTT TTATAGCTAA CCTCAACACACA ACCESSA A	
TOURSCEAM GCANTITETA GANTANCTAN GCANTAGANA TYNCHCTICA ATGCAGANAG GCAGTATCTA CATGAGATTA TGANATTGCG GTTGCTTT	
CICITCACTE AAAAAATAA CTAAAACTCT AACTITCAGA AAAAATGATT GTACATATAG AAAACCCAAA GCATCTAAAC AATTAAATA AATAAGTA	ra 15937
CHAGATTAC TOGATACAGA GTCAACATAC AAATATCAAT TGTATGTCTA TATACCAGCA ACGATTCAAA AATGATTTTT ATAATACCAT TAAAAATT ACCTTAGTA ATAAATGTCA GAAAGATCTC CAAGAACTCT ACATAAAAA TTATGAGACG TTATTGAGAA AAATTAAGGA AAACCTAAAT AAATGAAT	NG 16037
MAGGCANTG TITATCATTA AAGGATACAA TATAGTAAAT ATATCAAATG TITACTAATG GATTCAATGC AATACCAAAG TGCCAGCAGC CTITITTG	
GCTGGGGGGT CGGGCAGGAT TLATAAGCTA ATTATAAAAT GCATATGGAA ATGCAAAGGA CCAAGGATAG CCAAGACACT TTTCAGCAAC AATAAAAT	
TACTACITAC ACTACCAGAT GTCAAGACTT ATTATCGAGT TACATTTATT LAGACAGTGT GCTACTGACA CLAGGATAGA CAALGACAGA ACGACAAGA	
ACTAGAGIGE TEAGAAGEAE ACETGIACAT ATATAAAGGE TEGATTIATG ATAGAGETGE CAGTGEAGTA GAGAAGGAAA TTATTGETCH THE TAGAGAGA	
AGIVATAGG TUAATTAGAT ATTUATATGG CATGAAGTAT GARACAATAA CARTTATAT TUATARCTTG CAGARGCAR REPTORTA ARRESTA ARRESTA	
- AGTGATCACC ATAAAGGAAA AGATTGATAA ACTGGACTAT ATTAAAACTA AGGACTCCTG TYCAGCAAAA GACACTACTY CCACTGAAAA CACAACTA	
- IGAGTGAGAC AAGATATETG CAATACAGAT ACCTAATAAC TGAACCCAT ACACCAGATGATGA TGCGAATTA ACTTCCTACA ATCATTTAC AAAACCCA	
GCAGTATCT ACTAGATCTG AACATGTGAT CCAGTAATTA CACTCATAAT TATAAGCCAG TAAAAAGGCA TGTTTATGTC ACCALAAGAT ATATACA	F1 75037
INSTITUTE CACTATTATA CATALGAGCO AAAAACTGGA AACAAACCAA ATATCCATTA ACAGTAGAAT GAATALATAA AAGCTGTAAT AGTAATAC	NG 17037
TOGATACTA CACAGCAATG TAAATGAACT ACTGCTGTAC AAAACAACAT GGTTTAATCT CACAGACAAA ATGTTAAATG AAAGACACAG ACGAGTAC	AT 17137
ATTECGAACT TETGTTTATA ATTEMAGANE TEGCAAGANE TETTTACTET CITAGAAGTE CAGGTAATEG TANCETATAA AMAGGMAMA GEGTEGAA ATTEGGAEGE GECATETTET GEGETETTEA TANTETECTA TETATTEGTE AGTTTAGTET TTANACAGGE TEATTTACTT TETGAAAACT TACACTMA	TG 17237
HONGTOTAT TITTIGAATA TATGITATAC ATTAATAAAT AGGGTTTITA AACCIGTAGT TCATAATITA GGGAAAGTAG AATAACCAAA CATTAATA	AA 17337
IMACCANTE ANTINIAGIG CINCENTENT ITTINIGENT INTEGRANG TITATITIAE CITTETITEC NETETIATIT CANGGETELA ANATHRES	CP 17877
CCCAACGTA TATTGGGGCC AACATGAATG CCCCCAATGT ATATTTGACC CATACATGAG TCAGTAGTTC CATGTACTTT TTAGAAATGC ATGTTAA	TG 17637
. 85	
ASP VAI THE CYS ASH ILE LYS ASH GIY ARG CYS GIU GIN PHE CYS LYS ASH SET A ANCESTIAC TOTCHITTE GCTTCTTTE GAI GEA ACA TOT AAC ATT AAG AAT OGC AGA TOC GAG CAG TIT TOT AAA AAT AGT O	la
AGETGIAC TGICIATITI GCTICITITA G AI GTA ACA TGI AAC ATI AAG AAT OGC AGA TGC GAG CAG TIT TGI AAA AAT AGI G	CT 17724
MP AST LYB Val Val Cys Ser Cys Thr Glu Gly Tyr Arg Leu Als Glu Abn Gln Lys Ber Cys Glu Pro Als GNT ANC ANG GTG GTT TGC TCC TGT ACT GNG GGA TAT CGA CTT GCA GAA ANC CNG ANG TCC TGT GNA CCA GCA GC GTCATAN	
WE ARE DEED OF THE THE THE THE DAG GON THE CON CITY GON AND THE THE THAT GON GON GON GONTANT	CT 17807
CANALGATE TITIAAAGAA AATCIGTAIC IGAAACIICA GCAITITAAC AAACCIACAI AATTITAATI CCIACTIGAA ICTGCIICCI TITGAAA	
	CA 17967
INGAAAATAT CAGTAGCTTG AATTAGACCA ATTAATTTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATCT ACAACCTGAA TTCTTTCT	GT 18007
INGANATAT CAGTAGCTTG AATTAGACCA ATTAATTTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATCT ACAACCTGAA TTCTTTCI GTCAATTTG TCCAATTTT TTCTCTAACA TTTATATCAC AAAGCAATTA ATTTGTGTGA TTTCTGCATA TGTATTTGTA ATTCATCAG TCAATCA	GT 18007
INGALATAT CAGTAGCTTG AATTAGACCA ATTAATTTTC TAGATTGCAI CATATTTTAA ATATAACTAT GTAATCATCT ACAACCIGAA ITCTITCI STCAATTTT TCCTAATITTT TTCTCTAACA ITTATATCAC AAAGCAATTA ATTTGTGTG TTTTTGGATA TGTATTTGTA ATTCATCAG TCAAATCI STAGTAATAC TATATCATAA AATATACACA AATAATTGAG TGATAGGCTT CTAGTATAAG GACAGTAAGT TTGAAGCATG ATTCTATCTC COCTOCT	GT 18007 AT 18107 AG 18207
INGAMATAT CAGTAGCTTG AATTAGACCA ATTAATTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATCT ACAACCTGAA TTCTTTCT GTCAATTTG TCCAATTTT TTCTCTAACA TTTAATCAC AAAGCAATTA ATTTGTGTGA TTTCTGCATA TGTATTTGTA ATTCATCAG TCAAATCA GTAGTAATAC TATATCATAA AATATACACA AATAATTGAG TGATAGGCTT CTAGTATAAG GACGGTAAGT TTGAAGCATG ATTCTATCTC GGCTGGCT TTACTCTGA GAAAGTTATT TTTTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCAATG AACTGTTTA TGTTCTGC	GT 18007 AT 18107 AG 18207 TA 18307
INGAMATAT CAGTAGGTTG AATTAGACCA ATTAATTITC TAGATTGCAT CATATTITAA ATATAACTAT GTAATCATCT ACAACCTGAA ITCTITCI GTCLANTITIG TCCCAATTITI TICTCIAACA ITTATATCAC AAAGCAATTA ATTITGTGTGA TITCTGCATA TGTATTTGTA ATTCATCAG TCAAATCI GTMGTAATAC TATATCATAA AATATCACA AATAATTGAG TGATAGGCTT CTAGTATAAG GACGGTAACT TTGAAGCATG ATTCTATCTC GGCTGCCI HTACCTCTGA GAAAGTTATI TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCAATG AACTGTTTTA TGTTCTCG GGTGATCAG CACAATCTAT ATGGCTGGA ACAAACCAAT GTTTCCAGTC CATAGCACCAT TTTAACAGCT GATAGCTTA TTCAGAAC	GT 18007 AT 18107 AG 18207 TA 18307 AT 18407
INGALMATAT CAGTAGCTTG AATTAGACCA ATTAATTTTC TAGATTGCAT CATATTTTAA ATATAACTAT GTAATCATCT ACAACCIGAA ITCTITCT GTCAATTTT TCCTAATTT TTCTCTAACA TITATATCAC AAAGCAATTA ATTTGTGTA TITCTGCATA TGTATTTGTA ATTCATCAG TCAAATCT GTAGTATAC TATATCATAA AATATACACA AATAATTGAG TGAIAGGCTT CTAGTATAAG GACGGTAAGT TTGAAGGCATG ATTCTATCTG GGCTGCT HIACTCTGA GAAAGTTATT TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCAATG AACTGTTTTA TGTTCTGC GGGGTCTGATCAG CACAACCTAT ATGGCTGTGA ACAAAACAAT GTTTCCCAGT CATACCAACC ATGCCACCAT TITAACGGCT GATTAGTGTTA TCCAGAAC GTCGATCCCA TGTTCGTATG GCTGTTATCT AAAGATGAAA GCAGTGGACA CTTTTATTTT TTGALAALTT TAGGCTTCTG AGGGTCAALTT ATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	GT 18007 AT 18107 AG 18207 TA 18307 AT 18407
INGALMATAT CAGTAGCTTG AATTAGACCA ATTAATTITC TAGATTGCAT CATATITTAA ATATAACTAT GTAATCATCT ACAACCIGAA ITCITTCE STCAMITIT TCCAMITIT TICTCTAACA ITTATATCAC AAAGCAATTA ATTATGTGA TICTGGATA TGTATTIGTA ATTCATCAG TCAMATCA STAGTAATAC TATATCATAA AATATACACA AATAATTGAG TGAIAGGCTT CTAGTATAAG GACGGTAAGT TTGAAGCATG ATTCATCTG GGCTGGCT STACTCTGA GAAAGTTAIT TITTATTGTT GGGTCTTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCCAATG AACTGTTTA TOTTCTGG GGTGATCAG CACAATCTAT ATGGCTGGTA ACAAAACAAT GTTTCCCAGT CATACCAACC ATGCCACCAT TITTAACAGCT GATTAGTGTA TTCAGAAC STOCKCTCCA TGTTCGTATG GCTGTTATCT AAAAACTAATA GTCTCATTTATTTTTTTTTT	GT 18007 AT 18107 AG 18207 TA 18307 AT 18407 TA 18507 CT 18607
INGALMATAT CAGTAGCTTG AATTAGACCA ATTAATTTTC TAGATTGCAT CATATTTAA ATATAACTAT GTAATCATCT ACAACCIGAA ITCTITCT GTCAATTTT TCCCAATTTT TTCTCTAACA TITATATCAC AAAGCAATTA ATTTGGTGA TTCTGGATA TGTATTTGTA ATTCATCAG TCAAATCA GTGATTAC TATATCATAA AATATACACA AATAATTGAG TGAAGGCTT CTAGTATAAG GACGGTAAGT TTGAAGGCATG ATTCTATCTC GGCTGGCT TITACTCTGA GAAAGGTTATT TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCCAATG AACTGTTTA TGTCTCGG GTGATCAG CACAATCATA ATGCCTGTGA ACAAAACAAT GTTTCCCAGT CATACCAACC ATGCCACCAT TITAACGCT GATAGGTGAT ATATTTG LITCAGCGGGC TITTTTGAAG CACACTGAT ATAATTTCTT TTGCAATTCT AAAGCCTGAT ATCTTATTAA TTGGTACATT AAATTGTCCA CACTTCCT GTAACTGTT CAGTACCTGT CACACCAGA ATAATTTCTT TTGCAATTCTT AAAGCCCAGCT ACCTGCCTGA CACCTCCGAA CACCCAGACC GTAAGGGGA TATAATACACAC ACACACAGA ATAATAGAGGAA AGAAGAACCA GTGCCAGACT AGCTGGCTA GAGGCCCGAA AACCAGGAA ALTAGAGGAA TAAAAGACAA ACACACAGAA ATAAAGAGAACCA CTGCCAGACCT TCACAGCCCT AGCTGCCGAA CACCCCGAA CACCAGAGACCCTA AACCAGGAA ALTAGAGGAA TAAAAGACAA CACACAGAA ATATAGAGGAA TAGAGAGGAA CACCAGGCCCTA CACCAGGACCCCAA CACCAGGAA ALTAGAGGAA TAAAAGACAA CACACAGAA ATATAGAGGAA TAGAGGAGA CACCAGGCCCTA CACCAGGAA CACCAGGAA CACCAGGAA AACCAGGAA CACCAGGAA CACCAGGAA CACCAGGAA CACCAGGAA CACCAGGAA CACCAGGAA CACCAGGAA CACCAGGAACCAGAA CACCAGGAACCAGAA CACCAGGAA CACCAGGAACCAGAA CACCAGGAACCAGAA CACCAGGAACCAGAA AACCAGGAACCAGAA CACCAGGAACCAGAACCAGAA CACCAGGAACCAGAACAACAACAACAGAA AACCAGAGAACCAA CACCAGGAACCAGAACCAGAACCAAGAACAACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAA	GT 18007 AT 18107 AG 18207 TA 18307 AT 18407 TA 18507 CT 18607 GC 18707
INGAMATAT CAGTAGGTTG AATTAGACCA ATTAATTITC TAGATTGCAT CATATITTAA ATATAACTAT GTAATCATCT ACAACCTGAA ITCATTCT GTCCANTITTG TCCCAATTITT TICTCCHACA ITTATATCAC AAAGCAATTA ATTTGGTGGA TITCTGCATA TGTATTTGTA ATTCATCAG TCAAATCI GTMGTAATAC TATATCATAA AATATCACA AATAATTGAG TGATAGGCTT CTAGTATAAG GACGGTAACT TTGAAGCATG ATTCTATCTC GGCTGGCT HTACTCTGA GAAAGTTATT TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGGCAATG AACTGTTTTA TGTCTTGC GGTGGTCAG CACACCTATA ATGGCTGGTAA AGAAACAAH GTTTCCCAGT CATACCAACC ATGCCACCAT TITTAACAGGCT GATAGGTGTA TCAGAAC GTCCACTCCA TGTTCGTATG GCTGTTATCT AAAGATGAAA GCAGTAGACA CTTTTATTTT TTGAAAAATT TAGGCTCTGC AGGGTCAATT ATATTTGI HTACGGGGC TITTTTGAAG CAAACTAGAT ATAATTTCTT TTGCATTTCT AAAAGAACAA GTTCTATTAA TTGCTACATT AAATTGTCACACCA CACTTCTC GTACCTCT CACTACCTCT CTCAGCACTA ATAATTTCTT TTGCATTTCA AAAAGAACCA GTGCCAGATC AGCTTCGTCA GGAGCCCTA ATCCTGCC ACTAGGGGA TTAAAGACAC ACACCAGAA ATATAGAGTA TGCGAGGTGGGA AATCAGGGGT CTCACAGCCT TAGGACCTGA GAGGCCCGAA CAGAGGATT ACTAGGGGA TTAAGACACC ACACCAGGAA ATATAGAGTTA TGCGAGGTGGGA AATCAGGGGT TCACAGGCT TCAGACCTGA GAGGCCCCGAA CAGAGGATT ACTAGACGAA TTAAGACACC ACACCAGGAA ATATAGAGTTA TGCGAGGTGGGA AATCAGGGGT CTCACAGCCT TCAGACCTGA GAGGCCCCGAA CAGAGGATT ACCTAGCTATT TATTGACAGC AAGCCAGGAA ATATAGAGTTA TGCGAGATTAT CCTTATGGGCA AATAAAGGGA TGCACGCTGA TATTGTTCACACCTGCTAATTTCTACACACCTGATTATTCACACCTGCTAATTTCACCACCTGATTATTCACACCCTGATTATTATTCACACCCTGATTATTACACACCTGATTATTCACACCCTGATTATTACACACCTGATTACCCTGATTATTACACACCTGATTATTACACACCTGATTATTACACACCTGATTACCCTGATTACCCTGATTACCACCTGATTACCACCCTGATATACCACCAC	GT 18007 AT 18107 AG 18207 TA 18307 AT 18407 AT 18507 CT 18607 GC 18707 TA 18807
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTITC TAGATTGCA CATATITAA ATATAACTAT GTAATCATCT ACAACCTGAA ITCITTCI FICKANTITG TCCRAFTITI TICTCIACA TITATATCAC AAAGCAATTA ATTTGTGTGA TITCTGCATA TGTATTTGTA ATTCATCAG TCAAACCT FINGTMATAC TATATCATAA AATATCACA AATAATTGAG TGAATGGCTT CTAGTATAAG GACGGTAAGT TIGAAGCATG ATTCATATCC GCGTGGCT HIACTCTGA GAAAGTTATI TITTATTGTI GGGTCTTAAG CTGAGTTTAA CACATTGGTG TCAGAAAGAAT ATCTGTTTA TGTTCTGC GCTGATCAG CACAACCTAT ATGGCTGGTA ACAAAACAAH GTTTCCCAGT CATACCAACC ATGCCACCAT TITAACAGCT GATTAGCTA TCAGAAC CTCCACTCCA TGTTCGTATG GCTGTTATCT AAAGATGAAA GCAGTAGACA CTTTTATTTT TTCAAAAATT TAGGCTCTGC AGGGTCAATT ATATTTGI HIGAGGGGC TITTTGAAG CAAACTAGAT ATATTTCTI TTCCATTTCT AAAGCCTGAT ATCTTATTAA TTGGTACCATT AAAATTGTGCA CCATTTCT GTAACTGTT CAGTACCTGT CTCAGCACTA TACCAGGGCA AAGAATAAA GAAAGAACCA GTGCCCAGACT AGCTTGGTCA GAGGCCCGAA CAGAGAT ACTAGGGGAA TTAAAGACAC ACACAGGAA ATATAGAGTA TGGACTGGGA AATCAGGGGT CTCCACAGCCT TCAGAGCTGA CAGGTTACTGC CAGAGATT ACTAGGATA TATAAGACAC AACACAGAA ATATAGAGTA TGGACTGGGA AATCAGGGGA TGAGTTGGTCC TAGGACTC ACTAGCTTA AGGCCAGATA ACCACCAGAA ATATAGAGTA TCGAACTTAT CCTTTAGGGA AATAAAAGGGA TGAGTCTTGGC TAGGTATCTGC CAGAGATC ACCACAATTA ACTTTAGCAA TTGTCTGTCG TTAAAGAACCA CCTTTAAGCA GTTTTCCGCC CGGGTGGGC CAGGTGTTCC TTCCCCTI	GT 18007 AT 18107 AG 18207 AT 18307 AT 18407 AT 18507 CT 18607 GC 18707 TA 18807 AC 18907 AT 19007
INGALMITAT CAGTAGCTG AATTAGACCA ATTAATTITC TAGATTGCAT CATATITIAA ATATAACTAT GTAATCATCT ACAACCIGAA ITCITTCT STCAATTIT TCCTATTIT TICTCTAACA ITTATATCAC AAAGCAATTA ATTTGTGTGA ITCITTGGATA TGTATTGTA ATTCATCAG TCAAATCA GTAGTAATAC TATATCATAA AATATACACA AATAATTGAG TGAATGAGCTT CTAGTATAAG GACGGTAAGT TTGAAGCATG ATTCTATCTC GCCTGGCT TITACTCTGA GAAAGTTATT TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCCAATG AACTGTTTA TGTCTCGC GCCGATCGA GAACCTAT ATGGCTGTGA ACAAAACAAT GTTTCCCCAGT CATACCAACC ATGCCCACAT TTTAACGCT GATTAGTGTA TCCGGAATA ATATTTGT HICAGGGGC TITTTTGAAG CAAACTAGAT ATAATTTCTT TTGCAATTCT AAAGCCTGAT ATCTTATTAA TTGGTACATT AAATTGTGCA CCATTCC GTACCTCTT CAGTACCTGT CTCAGCACTA TACATTCTT TTGCAATTATA TTGGTACATT AAATTGTGCA CCATTCC GTACCTGTT CAGTACCTGT CTCAGCACTA TACATTCTT TTGCAATTATA TAGGCTCTGC AGGCTGTA GCCTGCCGAA CAGAGACCTA ATCTGGCC AAGAGGGAA TTAAAGACAC ACACACAGAA ATATAGAGTA TGGAGTGGGA AATCAGGGGA TCAAGCCTT CAGACCTGA GAGCCCGAA CAGAGAGTA CCACATATT TATTGACAGC AAGCCAGCA TAAGATTTAC TGAAAGTATT CCTTATGGGA AATAAAGGGA TGAGTCTGGC TAGTTATCTG CAGCAGGG HIGTCCTTAA GGCCAAAACT ACTTATGGGA TAACAAGGCA TACAGAGCA TATAAAGCA GTTTCCGCC CTGGCTGGGC CAGGTGTTCC TGCCCAA CCGCAAACCT CCACAACCT CCACTGGGGA TACACAGGA TATAAAGCA CCTTTAAGCA GTTTCCGCC CTGGCTGGGC CAGGTGTTCC TGCCCAA	GT 18007 AT 18107 AG 18207 TA 18307 AT 18407 TA 18507 CT 18607 GC 18707 TA 18807 AC 18907 AT 19007
INGALATRA CAGTAGGTG AATTAGACCA ATTAATTITC TAGATTAGAT CATATITAA ATATAACTAT GTAATCATCT ACAACCTGAA ITCTITCI FICENATITIG TCCCAATTITI TICTCCHACA ITTATATACA AAAGCAATTA ATTITGTGTGA TITCTGCATA TGTATTTGA ATTCATCAG TCAAATCI FINGTAATAC TATATCATAA AATATCACA AATAATTGAG TGAAAGCCTT CTAGTATAAG GACGGTAACT TIGAAGCATG ATTCTATCTC GOCTGGCT HITACTGTGA GAAAGTTATI TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCAATG AACTGTTTTA TGTCTTGC GGTGATCAG CACAACTATA ATGGCTGGTAA GAAAACAAH GTTTCCCAGT CATACCAACC ATGCCACCAT TITAACAGCT GATAGGTTA TCAGAAC GTGCACTCCA TGTTCGTATG GCTGTTATCT AAAGATGAAA GCAGTAGACA CTTTTATTTT TTGAAAAATT TAGGCTCTGC AGGGTCAATT ATATTTGI HICAGGGGC TITTTTGAAG CAAACTAGAT ATAATTTCTT TTGCATTTCT AAAAGAACAA GTTCATTATTAA TTGCTTCACTT AAAATTGTCAC CCATTTCC HICAGCGGC TITTTGAAG CAAACCAGAA ATAATAGAGTA TGCAGTAGGA AAACAAGAA ATAATGTCAC AAGAGTAGAACA GTGCCAGGAT AGCTTGGTCA GGAGCCCTA ATCCTGCC ACAGAGGAATAT TATAGACAGC AAGCCAGAAA ATATAGAGTA TGGAGTGGGA AATAAAGAGA GTTCAGGCG TAGTAGACTAG CAGCAGGAATCAAGAATCAA GAAAGAACCA CTCTAGGCA AATAAAGGA TGGAGTGTGGC TAGTATTCAG CACCAGGAA HIGGCCAGAT TTATGACAGG AAGCCAGGAA TTACATGTGG TTAAAGAACA CCTTTAAGCA GTTTTCCGCC CTGGGTGGGC CAGGTGTTCC TTGCCCCC ATGGCTAAC CCACAACCTA CAGATTAGCAA TTGCTCTGTGG TTTAAGAACA CCTTTAAGCAA GTTTTCCGCC CTGGGTGGGC CAGGTGTTCC TTGCCCCC ATGGCTAAAC CCACAACCTA CAGATTAGACAA TTGCTCTGTGG TTTAAGAACA CCTTTAAGCAA GTTTTCCGCC CTGGGTGGGC CAGGTTTTC CCAGGTTTC CTGCCCCC ATGGCCAGAT TTGCTAGCAA TTGCCCAACAACCA TTCCCAACCT TATCACACTG CTCCCAACACTT TTGCTTTATGG CCAGGTTTTG CCTGCCCCCAACACCT CCAGGTTTTG CCTGCCCCCAACACCT CTGCCCCCAACACCT CTCCCCAACAACACACACACACACACA	GT 18007 AT 18107 AG 18207 TA 18307 AT 18407 TA 18507 CT 18507 GC 18707 TA 18907 AT 19007 TT 19107 CT 19207
INGALATITA CAGTAGCTG AATTAGACCA ATTAATTITC TAGATTGCA CATATITAA ATATAACTAT GTAATCATCT ACAACCTGAA ITCITTCI FICKANTITG TCCAATTITI TICTCHACA TITATATCAC AAAGCAATTA ATTTGTGTGA TITCTGCATA TGTATTTGTA ATTCATCAG TCAAATCA FINGTAATAC TATATCATAA AATATACACA AATAATTGAG TGAATGGCTT CTAGTATAAG GACGGTAAGT TIGAAGCATG ATTCATATCG GCGTGGCT HIACTGTAG GAAAGTTATI TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCAATG AACTGTTTA TGTTCTGG GCTGATCAG CACAACCTAT ATGGCTGGTAA AAAAACAAA GCAGTAGACA CTTTAATTTT TIGAAAAATT TAGGCTCTGC AGGGTCAATT ATATTTG HIGAGGGGC TITTTTGAAG CAAACTAGAT AAAACTAAA GCAGTAGACA CTTTAATTTT TIGAAAAATT TAGGCTCTGC AGGGTCAATT ATATTTG HIGAGGGGC TITTTTGAAG CAAACTAGAT AAAACTAATAA GAAAGAACCA GTGCCCAGATC AGCTTGGTCA GGAGCCCCTA ATCTGGC GTAACCTTT CTAGACCACAT ATACTAGGAGA AGAAATAAA GAAAGAACCA GTGCCCAGATC AGCTTGGTCA GGAGCCCCTA ATCTGGC GTAACCTTT CTAGAGCACA ATACTAGGAGA AGAAATAAA GAAAGAACCA GTGCCCAGATC AGCTTGGTCA GGAGCCCCTA ATCTGGC ACTAGAGTT TATAAGACAC ACACAGAA ATATAAGAGTA TGGACTGGGA AATCAAGGGGT CTCCACAGCCT TCAGAGCTGA GAGCCCCGAA CAGAGATT ACTCCTTAA GGCACAAATC ACCACAGCA TAACATTAC TGAAACTATT CCTTATAGGA ATATAAGGGA TGAGTCTTGGC TAGGTTCC TTCCCCTT ATGGTAAAC CCACAACCTT CCAGTGGGA TATCAAGGGC ATCACAGGGA TATCAAGGGA TTTCCTTTATG GCCAGTTCC TCCCCCT ATGGTAAAC CCACAACCTT CCAGTGGGA TATCAAGGGC ATCACAGGGA TATCACAGTG CTGCCAGAGAT TTTCTTTATG GCCAGTTTC TCCCCCT ATGGCAAACT CCAGTGGGA TATCAAGGGC ATCACAGGGA TATCACAGTG CTGCAGAGAT TTTCTTTATG GCCAGTTTC TCCCCCT ATGGCAAACT CCAGTGGGA TATCAAGGCC ATCACAGAGC TATCACAGTG CTGCAGAGAT TTTCTTTATG GCCAGTTTCT GCCCTT ATGGCAAACT TCCGAGCCT CCAAACACAC AACCAGAAC TATCAAGGCC ATCACAGGGA TATCACAGTG CTGCAGAGAT TTTCTTTATG GCCAGTTTTTATAACACAC CCTTGTGTATATTTTATATG GCCAGTTTTTTATATATATATTTTATATATATATATTTTATATAT	GT 18007 AT 18107 AG 18207 TA 18307 AT 18407 TT 18607 GC 18707 TA 18807 AT 19007 TT 19107 CT 19207 CT 19307
INGALATITA CAGTAGETTG AATTAGACCA ATTAATTITC TAGATTGCAI CATATITTAA ATATAACTAT GTAATCATCT ACAACCIGAA ITCITTCI GTCAATTIT TCCCAATTIT TICTCTAACA ITTATATCAC AAAGCAATTA ATTTGTGTGA TITCTGGATA TGTATTGTA ATTCATCAG TCAAATCA GTCAATTAC TATATCATAA AATATACACA AATAATTGAG TGAATGAGCT CTAGTATAAG GACGGTAAGT TGGAAGGAT ATTCATCAG GCGGGGGT TITACTCTGA GAAAGTTATT TITTATTGTT GGGTCTTAAG CTGGACTTAC ACACTTGGGT TCAGAAGGAT TCCGGCCAATG AATCTTTTA TGTTCTGC GCCGATCAG CACACCTAT ATGGCTGTGA ACAAAACAAT GTTTCCCAGT CATACCACC ATGCCCACCT TTTAACAGCT GATTAGTGTAT TCCGGAATGAT ATTATTTT HIGAGGGGGC TITTTTGAAG CAAACTAGAT ATAATTTCTT TTGCATTCT AAAGCCTGAT ATCTTATTAA TTGGTACATT AAATTGTGCA CCATTCC GTACCTGTT CAGTACCTGT CTCAGCACTA TACCAGGCGA AAGAAATAA GAAAGAACA GTTCCAGCCT TCAGAGCCTGA AGGCCCCGAA CAGGAGCCCTA ACCTGGCC GTAAGGGGA TTAAGAGCAC ACACCAGGA ATATAGAGTAT TCGGAGGGA AATCAGGGGA TTAAGACCA CACACCAGA ATATAGAGTAT CCTTATAGGCA CTCACAGGCT TCAGAGCCTGA GAGCCCCGAA CAGGAGAT CCACATATT TATTGACAGC AAGCCAGGAA ATATAGAGTATC CGAAAGTATT CCTTATAGGCA GTTTCCGGC TCAGAGCCTGA GAGCCCCCGAA CAGGAGAT CCACATATT TATTGACAGC AAGCCAGGAA ATATAGAGTAT CCTTATAGGCA GTTTCCGGC CTGGGGGGGC CAGGTGTCC TTGGGAAAC CCACAACCT CCAGTATGGG TTTAAGAGC CTTTAAGGCA GTTTCCGCC CTGGGGGGGC CGGGGGTGCC TTGCCCCAA ATGCGTAAG CCCACAACCT CCAGTGTGGA TATCATGGGC TTTAAGAACA CCTTTAAGCA GTTTCCGCC CTGGGTGGGC CAGGGGTTCC TTGCCCTA ACCCCCACACCTT CCAGCACCT CCAGTGTGGA TATCAAGGCC TAACACGAGC TATCACAGG CTTGCCAAACCT TCAGAGGGCC TTCCCAAACC TCCAGAGGC TATCACAGC TAACACAGAAC ATCACAGGC TACCCAAACCT CCAGGGGCC TTCCCCAAACC ATCACAGGC TACCCAGAGC TATCACAGG CTTGCCCAAACCT TCCAGGGGC TTCCCCAAACC TCCAGAGGC TATCACAGC TAACACAGAAC ATCACAGAG CTTCCAAACC TACCCAGAGC TACCTCAAAGC TACCACACCT TCAGAGGGC TACCTACAGC TACCCCACACCTT CACATACACC TACCCAGAGC TACCCAGAGC TACCTCAAACC TACCCACACCTT CACATACACC TACCCAGAGC TACCCCACACCTT CACATACACC TCCCCAAAACC AACCACACCT TCCCCAAAACC AACCACACCT TCCCAAAACC TACCACACCTT TACCAGGC TACCACACCTT CACATACACC TACCACACCTT CAAGGTATATAT TATCCTGCAAACCTTA AAGAGAAAAA AACCAGAACC TACCACACCTT CAAGGTATATATA TATCCTGCAAAACATA AAGAGAAAACA GAATCAGAG CTACACACCTTA CAAGTTCTC TACC	GT 18007 AT 18107 AT 18107 AT 18207 TA 18307 AT 18407 CT 18607 GC 18707 AT 18807 AC 18907 AT 19007 TT 19107 CT 19207 CT 19307 AT 19407
INGALATITA CAGTAGGTTG AATTAGACCA ATTAATTITC TAGATTGCAT CATATITATA ATATAACTAT GTAATCATCT ACAACCTGAA TICTITCE FICENITITG TCCCANTITIT TICTCHACA ITTATATACA AAAGCAATTA ATTTGTGTGA TITCTGCATA TGTATTTGAA ATTCATCAG TCAAATCL FINGTAATAC TATACCATA AATATCACA AATAATTGG TGAAGGTTA CACATTGGTG TAGAAAGAT TIGAAGCATG ATTCATCACC FINGTAATAC GAAACTATA TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAAAGAT TCGGGCAATG AATCTTTTA TGTTCTGC GCTGATCAG CACAACTATA ATGGCTGGTGA ACAAACAAA GATTGTCTCCAGT CATACCACC ATGCCACCCAT TITAACAGGT CATATGGTAA TCAGAAG CTCCACTCCA TGTTCGTATG GCTGTTATCT AAAGATGAAA GCAGTAGACA CATTTCTT TTGAATATTT TTGAAAAATT TAGGCCTCGC AGGGTCAATT AAATTGCACACC MAGAGGGGC TITTTTGAAG CAAACTAGAT ATAATTCTT TTGCATTCTT AAAGCCAGAT AACTTATAAA TTGGTACATT AAATTGCACACCAAACTAAT TAACAGGGAGA AAACAAAAAAAA GAAAGAACAA AAAACAGAAAAAAAA	GT 18007 AT 18107 AG 18207 AT 18307 AT 18407 AT 18507 CT 18607 AT 18807 AT 18907 AT 19907 AT 19107 AT 19107 AT 19107 AT 19407 AT 19507
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATTGCAT CALATITAA ATALACTAT GTAATCATC ACAACCTGA ITCATTCT STCANTITIG TCCAATTTI TICTCHACA ITTAATACAC AAAGCAATA ATTGGTGA TITTGGAGA TGTATTGTA ATTCATCAG TCAAACCT STMGTAATAC TATACCATA AATATCACA AATAATTGG TGAAGGCTT CTAGTATAG GACGGTAACT TIGAAGCATG ATTCATCATC GACGGGC TITTACTCTAG GAAAGTTAT TITTATTGTT GGGTCTTAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCGGGCATG ATTCATCATC GACTGCTTATACAGGCT GATAGGTAA TCAGAAGACA CTCCACCCAT TITTAACAGGCT GATAGGTAA TCAGAAGACA GTCCACCAT TATAACAGGC GATAGGTAA TACATTGAAA STMCAGGGGC TITTTGAAG CAAACTAGA AAAATAAA GACAGAGAA CTTTTATTTT TIGAAAAATT TAGGCCCGC AGGGTCAATT AAATTGCTA MICAGGGGC TITTTGAAG CAAACTAGA AAAATTATCTT TTGCATTCCT AAAAGCCAGA ATCAGGGA TACATGAGA ATCAGGGA TACAGAGGA TACACAGGA TACAGAGGA TACACAGGA TACAGAGGA TACACAGGA TACAGAGGA TACAC	GT 18007 AG 18207 TA 18307 TA 18307 TA 18507 CT 18607 TA 18507 TA 18507 TA 18507 TA 18507 TA 18907 TA 19907 TA 19107 TT 19107 TT 19107 TT 19107 TT 19507 AG 19507 AG 19507 TA 19507 TA 19507
INGAMATAT CAGTAGCTG AATTAGACCA ATTAATTTC TAGATTGCAT CALATITAA ATALACTAT GTAATCATC ACAACCTGA ITCATTCT STCANTITIG TCCAATTTI TICTCHACA ITTAATACAC AAAGCAATA ATTGGTGA TITTGGAGA TGTATTGTA ATTCATCAG TCAAACCT STMGTAATAC TATACCATA AATATCACA AATAATTGG TGAAGGCTT CTAGTATAG GACGGTAACT TIGAAGCATG ATTCATCATC GACGGGC TITTACTCTAG GAAAGTTAT TITTATTGTT GGGTCTTAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCGGGCATG ATTCATCATC GACTGCTTATACAGGCT GATAGGTAA TCAGAAGACA CTCCACCCAT TITTAACAGGCT GATAGGTAA TCAGAAGACA GTCCACCAT TATAACAGGC GATAGGTAA TACATTGAAA STMCAGGGGC TITTTGAAG CAAACTAGA AAAATAAA GACAGAGAA CTTTTATTTT TIGAAAAATT TAGGCCCGC AGGGTCAATT AAATTGCTA MICAGGGGC TITTTGAAG CAAACTAGA AAAATTATCTT TTGCATTCCT AAAAGCCAGA ATCAGGGA TACATGAGA ATCAGGGA TACAGAGGA TACACAGGA TACAGAGGA TACACAGGA TACAGAGGA TACACAGGA TACAGAGGA TACAC	GT 18007 AG 18207 TA 18307 TA 18307 TA 18507 CT 18607 TA 18507 TA 18507 TA 18507 TA 18507 TA 18907 TA 19907 TA 19107 TT 19107 TT 19107 TT 19107 TT 19507 AG 19507 AG 19507 TA 19507 TA 19507
INGALATITA CAGTAGCTG AATTAGACCA ATTAATTTC TAGATTGCA CALATITAA ATATACTAT GTAATCATC ACACCTGA ITCATTCA CALACTGA ATTATATCA CALAGOATA ATTTTGGTGA TITCTGGATA TGTATTGTA ATTCATCAG GALACTAG GALACTAG TAGATTAGA ATTATACACA ATTATACACACA CACTTGGTG TAGACACACA TTAACAGCA GATCTATTA TGTATCAGACA CACTTGGTG TAGACACACA TTAACAGCT GATTAGCATA TACAGGAC CACACCTATA ATTATACACAC TACACCACACT TATACAGCT GATTAGCATA TACAGGAC CACACCTACACACA TGTACCACAC TACACACACA TACAGGACA ACACAGAAA ATTATACACAC ATTACAGCA TACAGGACA ATTATATATA TAGACACCACACACACACACACACACACACACACACACAC	GT 18007 AT 18107 AT 18107 AT 18207 AT 18307 AT 18507 CT 18607 AC 18707 AC 18907 AC 18907 AC 19907 AC 19907 AT 19007 AT 19007 AT 19507 AT 19507 AA 19607 AA 19807
INGALATITA CAGTAGETG AATTAGACCA ATTAATTITE TAGATTGCT CATATTTAA ATATACTAT GTAATCATE ACAACCTGA TICTITCE FICENTITIG TCCANTITIT TICTCHACA ITTATATACA AAAGCAATTA ATTTGTGTGA TITTGGATA TGTATTGTA ATTCATCAG CATACTTAT FINGTAATAC TATACCATA AATATCACA AATAATTGG TGAAGCATT CTAGTATAG GACGGTAACT TIGAAGCATG ATTCATCAG CACTGCT TITACTCTAG GAAAGTTAT TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCGGGCAATG ATTCATCAC CACTGCTATTACAGGT GATACGTTA TATACAGGA CACTACTACACC ATTCACACCA TITTAACAGGT GATACGTTA TATACAGGA CTCCACTCCA TGTTCGTATG GCTGTATCT AAAGATGAAA GCAGTAGACA CACTTCCTA ATTCACAGCA TATACAGGAC CACTACCACCA TITTACAGGAC TATACTGTA AAACAGAAA GAAATGAAA GCAGTAGACA CACTTCCTA AAACCAGAA ATAATACACA AAACAACAA ATAATTCTT TTGCATTCT AAAACACACAA ATAATACACA AAACAACAA ATAATACACA AAACAACAA AAAAAAAA	GT 18007 AG 18207 AG 18207 TA 18307 TA 18507 CT 18607 TA 18507 TA 18507 TA 18507 TA 18507 TA 18507 TA 19007 TT 19107 TT 19507 AA 19407 TT 19507 AA 19507 AA 19807
INGALATITA CAGTAGCTG AATTAGACCA ATTAATTITC TAGATTGCA CALATITAA ATATAACTAT GTAATCATCT ACAACCTGAA TICTITCE FICKANTITIG TCCAATTITA TICTCAACA TITATATCAC AAAGCAATTA ATTTGTGTGA TITCTGCATA TGTATTTGTA ATTCATCAG TCAAACCTA FINGTAATAC TATATCATAA AATATACACA AATAATTGAG TGAATGGCTT CTAGTATAAG GACGGTAACT TIGAAGCATG ATTCATCACG FINGTAATAC GAAAGTATAT TITTATTGTT GGGTCTTAAG CTGAGTTTAC ACACTTGGTG TCAGAATGAT TCCGGCAATG ATTCATCACG GCCGATCTAC ATTCATCATA ATGGCTCGTGA AAAAACAAA GATTCCCAGT CATACCAACCA TGCCACCCAT TITAACAGCT GATTAGCTTA TCCAGAAC CTCCACTCCA TGTTCGTATG GCTGTTATCT AAAGATGAAA GCAGTAGCAC CTTTTATTTT TTGAAAAATTA TAGGCTCTGC GAGTGGTCAAT TATATTTG AAAGAGGGCC TITTTTGAAG CAAACTAGAT ATATTTCTT TTGCATTCT AAAGCCTGAT ATCTTATTAT TTGGATACATT AAATTGTGCA CCATTTCT AAAGAGGGCC TITTTTGAAG CAAACTAGAT ATATTTCTT TTGCATTCT AAAGCCTGAT ATCTTATAAA TTGGATCAGT AAAATTGAAC CTCACTCTA GGGTACCTTC CTCAGCACAT AAAACTAGAT AAAACAGAAA AAAACAAA GAAACAAAAAAAAAA	GT 18007 AG 18207 AG 18207 TA 18307 AT 18507 CT 18607 TA 18807 AC 18907 AT 1907 TT 19107 CT 19207 TT 19207 AG 19407 TT 19507 AA 19407 TT 19507 AA 19607 AA 19807
INGALATITA CAGTAGCTG ARTTAGACCA ATTARATTITC TAGATTGCAT CATATITIAA ATATACATA GENATCATCH ACACCIGAA ITCATACT GENATURITI TICTCAACA TITATATCAC AAAGCAATTA ATTITGTGTGA TITATGCATA TITATATCACA ATTARATCACA CACTTGGTG TAGACATACA TITACAGCAT ATTARATCACA ACCTTGTTA TITACAGCAT GATTAGCATA ATTARATCACA CACTTGGTG TAGACATACA TITACAGCAT GATTAGCATA ATTARATCACA CACTTGGTG TAGACATACA TITACAGCAT GATTAGCACA CACTTGGTG TAGACACCA TITACAGCT GATTAGCACA CACTTGGTG TAGACACCA TITACAGCT GATTAGCACA CACTTGGTG TAGACAGCA TITACAGCAT TAGACAGCA ATTARATCACA ATTARATCACA ATTARATCACA ATTARATCACA ATTARAGCAA	GT 18007 AT 18107 AT 18107 AT 18207 AT 18307 AT 18507 CT 18607 AT 19607 AT 19007 AT 19007 AT 19007 AT 19007 AT 19507 AT 20007 AT 20007
INGALATAT CAGTAGCTG ARTRACACA ATTANTTITC TAGATACAT ATTATATACA TATATACAT GTAATCAT ACAACTGA TICTITCE GTCANTITG TCCANTITT TICTCHACA ITTATATACA AAGGARIA ATTGGGGA TITTGGAIA GTATTGTA ATTCATCAG CAGACTGA GTAGATAC TATACATA AATATCACA AATATTGG TGAAGGCTT CTAGGATAGG TITTGGGGA TTGAAGGCATG ATTCATCAG CAGAGGTT GTAGATGG CACAACTAT ATGGCGGGA ACAAACAAG GTTCCCAGT CATACCAGC ATCCACCAT TITAACAGGT GATAGGTAT TICAGAAG CTCCACTCCA TGTTCGTATG GCTGTATCT AAGATGAAA GCAGTAGACA CTTTTATTT TIGAAAAATT TAGGCCTGC GATAGGTA TACAGAAG MICAGGGGC TITTTGAAG CAAACTAGA AAAATATAT TTGCATACAACC AGGGCCCATA TACAGAAGACA CTTTTATTT TIGAAAAATT TAGGCCTGC AGGGCCCATA TACAGAGAA MICAGGGGC TITTTGAAG CAAACTAGA AAAATATATCT TTGCATATCA AAAAGAACA GTGCCAGGT AGCTAGACTA AAAATTGCCAGA MICAGGGGC TITTTGAAG CAAACTAGA AAAATTATCT TTGCATTCCT AAAAGAACAA GTGCCAGGT AGCTAGACTA AAAATTGCCAGA MICAGGGGA TTAGAAGACA CACACAGGA ATATAGGGT TGCAGAGGGA AAAAAAAAA GAAAGAACAA ATATAGGATA TACAGGGGA AAAAAAAAAA	GT 18007 AT 18107 AT 18107 AT 18207 AT 18307 AT 18507 CT 18607 AT 19607 AT 19007 AT 19007 AT 19007 AT 19007 AT 19507 AT 20007 AT 20007
REGALATAT CAGATEGAS ATTACACA ATTACATOR CALASTICAL CATATITAL ATLACAT GRANTCAT CACACTGAL ITCOTTCE GREANTING TCCANTITIT TECTALACA ITLATATECA ANAGGANTA ATTIGGGA TITCHCATA GREANTAC TATATCATA ANATATCACA ANAGANTA ATTIGGGAT TITCHCATA GREANTAC TATATCATA ANATATCACA ANAGANTAC CAGAGGATAG TITCHCAGA GREANTAC CACATCATA ATGGCTOGGA ACAAACAAI CATTCCCGG CATACCAGG CAGAGGAT COCACTCCA RETICGATAG GCCGATTAC ANAGATGAN GENERACAG CACATCATA TAGGCTOGGA ACAAACAAI CATTCCCGG CATACCACC COCACTCCA RETICGATAG GCCGATACA ANAGATGAN GENERACAG CACACTGAT ATGCCTCAGAGAA CACACGAAA GAACTAGAT ATAATTICTI TICCATTCT ANAGCCAGAT ATCATATAA TICCGACAC AGGCCCCATA CACACAGAA ATAACACC ACACACAGA ATAACAGGA TACACAGGAA AACAAGGAA CACTTGAGGA ATCACAGCA CACACATAT TATTGACAGC AGCCCACTA TACACAGGAA ATAACAGAA CACACAGAA ATAACAGAA CACACAGAA ATAACAGAA CACACAGAA ATAACAGAA CACACAGAA ATAACAGAA CACACAGAA ATAACAGAA CACACAGAA ATACACAGGA TACACAGGAA CACACAGAA ATACACAGAA CACACAGAA TACACAGGAA CACACAGAA ATACACAGGA CACACAGAA TACACAGGAA CACACAGAA ATACACAGGA CACACAGAA TACACAGGAA CACACAGAA TACACAGGAA CACACAGAA TACACAGGAA CACACAGAA CACACAGA CACACAGAA CACACAGAA CACACAGAA CACACAGAA CACACAGAA CACACAGA	GT 18007 AT 18107 AT 18107 AT 18207 AT 18507 CT 18607 AT 18807 AT 18907 AT 19907 AT 19907 AT 19907 AT 19907 AT 19507 AT 20107
REGAMATAT CAGTRACTIC ANTRACECA ATTACTTIC TAGATTACAT CATATITAA ATATACTAT GRANTCATC ACACCIGAL TICCITCE STREATING CICANTITIT TICCTALACA TITATACEAC AAAGCANTIA ATTIGUEA TITCTGCATA TOTATTICAT ATTICATCAG TAANTLE STREATING CICANTITIT TICTTALCA AAACANTICA CICAGTTITA CACCIGGAT TITCAGCAT ATTICATCAG GOCTGGC TITACTCAG CACANTCATA TAGGCTGCTGA ACAAACANT CITCCCAGT CATACAACCA TAGCACCAT TITACAGAT ACGTTATTA TITCTCAGA CICACTCA TOTTCATAG GOCTGTAA AAAACANA CITCCAGT CATACAACCA TAGCACCAT TITACAGAT ACGTTATTA TITCTCAGAC CICACTCCA TOTTCAGAG CAAACTAGAT AAAATATA GAAGCAACA CITTATATTT TITCAGACAC TITCACAGAC AGGGCTCATATTATATA TAGGCTCTCA AGGGCTCATATTATATA MICAGGGGC TITTTGAAG CAAACTAGAT ATAATTCTT TITCATTCTT AAAAGATAAA GAAGAACAA CITTATTATT TITCATACATATAA TITCGTACATA AAAATATAA GAAGAACAA CITTATTATA TITCGTACATA AAAATATAA GAAGCACCATA TAAAATAAA GAAGCACCATA TACAACGAAA AAAAAAAAAA	GT 18007 AN 18307 AT 18407 AT 18507 AT 18507 AT 18507 AT 18507 AT 18507 AT 19507 AT 19907 AT 19107 AT 19107 AT 19507 AN
INGALATITE CONTACTION ANTIACACCA ATTANTITIC TAGATIGGAT CATATITATA TATALACTAT GRANICATC ACACCIGAA ITCITICA STAGIANTA TATACATAA AATATACACA AATAATIGGA TEATAGGCTT CAGATAGAG TITAGGCATA ISTATIGATA ATTACACAA STAGIANTAC TATACCATAA AATATACACA AATAATIGGA TEATAGGCTT CAGATAGAG TITAGAAGAT TITAGAGCATG ATTCATACT GEGGATCAG CACAATCTAT ATGGCTCOTGA ACAAACAATG CTAGATAGAC GECGATCAG CACAATCTAT ATGGCTCOTGA ACAAACAATG CTAGATAGAC CICACTCCA TEGTCGTATG GCTGTTATC AAGATGATA GCTGTTACACACA TICCACCACCAT TITAACACAT GAACCTGAT CICACTCCA TEGTCGTATG GCTGTTATC AAGATGAAA GCAGTAGACA CICACATCAC TITTTGAGA GCAAACTGAT ATAATTTCTT TICCATTCT AAAGCCTGAT ACCTGGCT MINAGGGGC TITTTGAAG CAAACTGAT ATAATTTCTT TICCATTCT AAAGCCTGAT ACCTGGCT ACCAACATT ATATGACAGC AACCACGAA ATAATTCTT TICCATTCT AAAGCACCAACACCAT TACCAGGCCAAACACACAA AACCACGAA ATAAACACCACACACAAACAAAAAAAAAA	GT 18007 AT 18107 AT 18107 AT 18207 AT 18507 AT 18507 AT 18607 AT 18607 AT 19007 AT 19107 CT 19207 AT 19207 AT 19507 AT 20107 AT 20107 AT 20107 AT 20107 AT 20307
INGALATAT CAGITAGCTIG ANTINGACCA ATTANTITIC TAGALIGGAI CALAITITIAA ATALACTAT GIAATCATT ACAACCIGAA ITCAITCE COCANTITIT TICTCAAACA TITAATCAC AAGCAMATA ATTINGTIGGA TITCACCAN GIATTIGGA ATTATGAA ATALACACA AATALTIGAG TGATAGGAT CIAGATAAG GACGGTAGT TIGAAGCATG ATTCATACAG CACAATCAT CIAGATAACA AATALTCAG TGATAGGAT CIAGATAAG TAGAACTAG TITAACACAG ATACACACC ATCACACCAC TITAACAGAT AATACACAA AATALTCAG CACAATCATA AATACACAA AATAATCAG CACAATCATA AATACACAA CACAACCAATCACACCAATCACACCAT TITAACAGCT GATAGGATA TAGACAGAT ACAATCATA AACAATGAAA GAACAACCAA GATACACACCA TAGACACCAA TACACACCAA TACACACCAA CACAACCAATCAA AAACAAGCAA GAACAACAAA AAACAAGGAAAAAAAA	GT 18007 AN 18107 AN 18207 AN 18307 AN 18507 CT 18607 AN 18907 AN 18907 AN 19907 CT 19307 AN 19407 AN 19507 AN
INGALATITE CONTACTION ANTIACACCA ATTANTITIC TAGATIGGAT CATATITATA TATALACTAT GRANICATC ACACCIGAA ITCITICA STAGIANTA TATACATAA AATATACACA AATAATIGGA TEATAGGCTT CAGATAGAG TITAGGCATA ISTATIGATA ATTACACAA STAGIANTAC TATACCATAA AATATACACA AATAATIGGA TEATAGGCTT CAGATAGAG TITAGAAGAT TITAGAGCATG ATTCATACT GEGGATCAG CACAATCTAT ATGGCTCOTGA ACAAACAATG CTAGATAGAC GECGATCAG CACAATCTAT ATGGCTCOTGA ACAAACAATG CTAGATAGAC CICACTCCA TEGTCGTATG GCTGTTATC AAGATGATA GCTGTTACACACA TICCACCACCAT TITAACACAT GAACCTGAT CICACTCCA TEGTCGTATG GCTGTTATC AAGATGAAA GCAGTAGACA CICACATCAC TITTTGAGA GCAAACTGAT ATAATTTCTT TICCATTCT AAAGCCTGAT ACCTGGCT MINAGGGGC TITTTGAAG CAAACTGAT ATAATTTCTT TICCATTCT AAAGCCTGAT ACCTGGCT ACCAACATT ATATGACAGC AACCACGAA ATAATTCTT TICCATTCT AAAGCACCAACACCAT TACCAGGCCAAACACACAA AACCACGAA ATAAACACCACACACAAACAAAAAAAAAA	GT 18007 AN 18107 AN 18207 AN 18307 AN 18507 CT 18607 AN 18907 AN 18907 AN 19907 CT 19307 AN 19407 AN 19507 AN
INGALATAT CAGTACCTTC ANTACACCA ATTANTITC TAGATECAT CATATITAA ATATACTAT GRATCATT ACACCIGAA TCATTTE STORMITTS TECRAACA ATATACACA ATATACTACA AACACAATA ATTITGETSAA TITCAGCAATA ATTITGATAA ATTITGATAA TITCAGCAATA ATATACACA AATATTCAC CAAGCAATA ATTITGATAAA TITCAGCAATG ATTICATACAC CACACATTTA TAGATTCAC CAGATACACA AATATTCAC CAGATACACA CACACATACACACACATA TAGATACACA AATATTCACACACACATA TAGATTCACACACACATA TAGATTCACACACACATA TAGATTCACACACACACATA TAGATTCACACACACACATA TAGATTCACACACACACATA AACACACACAAAACACACACAC	GT 18007 AN 18107 AN 18207 TA 18307 AT 18407 TA 18507 CT 18607 AN 18907 AN 19907 AN 19407 AN 19507 AN
IRGAMATA CAGTAGCTTG ANTRAGACA ATTANTITE TAGATTGCAT CATATTITAA ATATAACTAT GTAATCATC ACACCICAA TICTITCE STOCKHITTG TCCAATATA ARATACACA ARATATCACA CAAGGAATTA ATTITGGGGAT TICTIGGATA TOTATTGTAA ARATACACA ARATACACACA ATTITATCACA GAAGATTATA TICTICTACA GAAGATTATA TICTICTACA GAAGATTATA TITTATTGTT GGGTCTTAGA CACATTGGTG TCAGATAGAT TCCAGGCATA ACCTITATA TICTICTACACACACACACACACACACACACACACACACA	GT 18007 AN 18107 AN 18207 TA 18307 AT 18407 TA 18507 CT 18607 TA 18907 AN 19907 CT 19307 AN 19407 TO 19507 AN 19807 AN 19807 AN 19807 AN 19807 AN 19907 CT 20007 AN 20107 TA 20307 AN 20107 TA 20307
ISSAMATA CAGTAGCTIC ANTRACACA ATTANTITIC TAGATIGGAT CATATITIAA ATATAACTAT GATATCATC ACAACCIGAA ITCITICATION TOTACATATITI TICTCATACA ATATATCACA AAACACATAA ATTATGCATA TICTCACATATA GATATCATAA AATATACACA AATAATCACA CACAACTATA ATTATCACAGATCATA TICTCACATACATA TICTCACATACATA ACCACACATA CACCACTATA ATTATCACACA CACAACCATA AATATCACACACA	GT 18007 AN 18107 AN 18207 TA 18307 AT 18407 TA 18507 CT 18607 TA 18907 AN 19907 CT 19307 AN 19407 TO 19507 AN 19807 AN 19807 AN 19807 AN 19807 AN 19907 CT 20007 AN 20107 TA 20307 AN 20107 TA 20307
INGRANATE CAGTACCTE ANTRACACCA ATTANTITC TAGATTECH CATATITIAA ATATAACTAT GENATCATC ACAACCTGAA TECTTCE STOCKHITTE TICCTARACA ITTANTICA AMAGAATTA ATTAGATTAG CAGATTAATTAG CAGATTAATTAG CAGATTAATTAG CAGATTAATTAG TATAACTAA ATTAATTCAG TAGATAGAT STATICATAA ATTACATAA AATTAATCAG TAGATAGAT CAGTAGATGA TICCGCAATGA ATTCATCTC GENTATAG CACAACTATA ATGGCTGTGA AACAAAACAAT GITTACCAGC CATACCAACC ATGCCACCAT TITAACAGCT GATTAGTTTA TICCGCATT GENTATCAG CACAACTAGAT ATGGCTGTGA ACAAAACAAT GITTACCAGC CATACCAACC ATGCCACCAT TITAACAGCT GAGTAGATGAT TACCAGCAGCA STATICTAGAG CACAACTAGAT ATAATTCAT TIGCATTTCA AACACCAGCA TITTATTTT TICGACTATTAA TAGGCTTGAT AAGATGAAA GAGTAGAGCA CACACCAGCATA ATAATTCAT TIGCATTTCA AACACCAGCA TACCAACCA AGGCACCAT ACCATTCATAA MAGAGGGC TITTATGAG CAACCACCAGA ATAAATTCAT TIGCATTTCA AAACACCAGCA TACCAACCA GAGCACCATAAAAAA GAAAAAAAAAA	GT 18007 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 19107 19107 19107 19107 19107 19107 19107 19107 19107 19107 19107 19207 19207 19207 19207 19207 19307 19307 19307 19307 19307
INGRANATE CAGTACCTE ANTRACACCA ATTANTITC TAGATTECH CATATITIAA ATATAACTAT GENATCATC ACAACCTGAA TECTTCE STOCKHITTE TICCTARACA ITTANTICA AMAGAATTA ATTAGATTAG CAGATTAATTAG CAGATTAATTAG CAGATTAATTAG CAGATTAATTAG TATAACTAA ATTAATTCAG TAGATAGAT STATICATAA ATTACATAA AATTAATCAG TAGATAGAT CAGTAGATGA TICCGCAATGA ATTCATCTC GENTATAG CACAACTATA ATGGCTGTGA AACAAAACAAT GITTACCAGC CATACCAACC ATGCCACCAT TITAACAGCT GATTAGTTTA TICCGCATT GENTATCAG CACAACTAGAT ATGGCTGTGA ACAAAACAAT GITTACCAGC CATACCAACC ATGCCACCAT TITAACAGCT GAGTAGATGAT TACCAGCAGCA STATICTAGAG CACAACTAGAT ATAATTCAT TIGCATTTCA AACACCAGCA TITTATTTT TICGACTATTAA TAGGCTTGAT AAGATGAAA GAGTAGAGCA CACACCAGCATA ATAATTCAT TIGCATTTCA AACACCAGCA TACCAACCA AGGCACCAT ACCATTCATAA MAGAGGGC TITTATGAG CAACCACCAGA ATAAATTCAT TIGCATTTCA AAACACCAGCA TACCAACCA GAGCACCATAAAAAA GAAAAAAAAAA	GT 18007 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 18107 19107 19107 19107 19107 19107 19107 19107 19107 19107 19107 19107 19207 19207 19207 19207 19207 19307 19307 19307 19307 19307
INGRALATAT CAGATACCTIC ANTENATUTE TAGATEGAT CATATITATA ATTAAACTAT GTAATCATC ACACCIGAA TICTITAE STORTHITE TICCASTITTE TICTCAMACA INTENATICAC ANACGANTA NITIGITIGA TICTGCATA TICATATATA ANTENATACA CAATATATA ANTENATACA CAATATATATA NITIGATICA GOCTEGAT STORTHACTAG GARANTIAT TITTATITATIC GOCTEGAT STORTHACTAG GARANTIAT ATGCTCTGA ACAMACAT CITATCACAC TAGATACAC ANGCACAT TITAACAGCT GATTAGTAT TOTACTAG GEGATACAC CACATATAT ATGCTCTGA ACAMACAT CITATCACACCAC ANGCACACAT TITAACAGCT GATTAGTAT TOTACAGAC COCATCAC TOTTCCATATG GOCTTATAC ANAGATAMA GATTAGACA CITATATTI TICATATATA TITCACAGCT MAGGGGGC TITTATGAAG CAAACAGAT ATANTICTI TICATTCT ANAGCCTGAT ANCTATATATA MAGGGGGC TITTATGAAG CAAACAGAT ATANTICTI TICATTCT ANAGCCTGAT ANCTATATATA MAGGGGGC TITTAGAAGCAC ACCACAGAM ATANAGACTAA GAAGAATAAA GAAAGAACA CITATATTATAA TICGTACATT ANATICTACA CACATATT TAATGACAC ACCACAGAM ATANAGACTAA TICATACTACT TICACATTCT MAGAGGGA TAMACACAC ACCACAGAM ATANAGATTAA GAAGATAAAA GAAGAGACA CITATAGTAA TATATTGA MAGAGGGA TAMACACAC ACCACAGAM ATANAGACTAA TICATAGATTA CATATACAGCA TICACAGCCT MAGAGGAA TAMACACAC ACCACAGAM ATANAGACTA TICATAGATA TICATAGAGCA TICACAGCCT MAGACACAT TATATACACA TATATACAT TICATAGATA CATATAGATTA CATATACACCACACCT MAGACACAT TICACAGCACAT ACATACACT MAGACCAGAT TICACAGCACA TATACAGATACACT MAGACCAGAT TICACAGCCAC TICACACCTA CACATACACACT MAGACCAGAT TICACAGCACACACCT MAGACCACATT CACATACACT MAGACCAGATTACACT MAGACCAGATTACACT MAGACCACATTACACT MAGACCACACCTA CACATACACAC MAGACCACCTA CACATACACAC MAGACCACCTA CACATACACACCT MAGACCACCTA CACATACACAC MAGACCACCTACACCTA CACATACACAC MAGACCACCTACACCTA CACATACACACCACCTA CACATACACCACCTACACCTA CACATACACCACCACCTA CACACCTAC MAGACCACCACCTACACCACCACCACCTA CACACCTACACCACCACCTA CACACCTACACCACCACCACCACCACCACCACCACCACCA	GT 18007 181
INGALATAT CAGTAGOTTE ANTRACACCA ATTANTITIC TAGATTIGAT CITATITAA ATTANACTAT GRANTCATOT ACAACCIGAA ITCTITCE COCCURTIVE TICCATATITY TICCATORIA TATATICATACA ATTANACTAT ATTANACTAT ATTAGATACA ATTANACTAT ATTAGATACA ATTANACTAT ATTANACTAT ATTANACTAT ATTAGATACA ATTANACTATA ATTAGATACA CACACTAGAT ACACTICGAT CACACTAGAT TACACACA TICCAGCAT ATTAGATACA TAGAGATAA TAGAGATAA TAGAGATAA ATTAGATACA CACACTAGAT ATTAGATACA CACACTAGAT ATTAGATACA CACACTAGAT ATTAGATACA CACACTAGAT ATTAGATACA ATTAGA	GT 18007 181
INGALATAT CAGATACTIC ANTRACACCA ATTANTITIC TAGATIGAT CITATITAA ATATAACTAT GRANICATET ACAACCIGAA ITCATTCA CAGATITITY TICCATITY TICCATATITY TICCATATITY TICCATATITY TICCATATITY TICCATATITY TICCATAGA ATTANTICA ATTANTCATA ACACTAGAT CAGATITATA TICCAGACA ATTANTCATA ATTANTATA ATTANTCATA ATTANTCATA ATTANTCATA ATTANTCATA ATTANTCATA ATTANTATA ATTANTCATA ATTANTC	GT 18007 181
REGIMENTA CAGAGGATE ANTAGACCA ANTANTITIC TAGATIGGAT CAIATITHA ATAINACTAI GIAATCATC ACAGCCIGAL ITCATICA FROMITIA TICCATACA TITATACACA ANTANTAGA ANTANTAGA ANTANTAGA TATATACACA BATATACACA ANTANTAGAGA TATATACACA ANTANTAGAGA TATATACACA ANTANTAGAGA TATATACACA CAGAGGATA ANTANTAGAGA ANTANTAGAGA CAGAGGATA ATTATACACA ANTANTAGAGA CAGAGGATA CAGAGGATA TAGAACATA TAGAGCAGA ANTANTAGAGA CAGAGGATA CACATAGAGA CACATAGATA TAGAGCAGA ANTANTAGAGA CACATAGAGA CACATAGAGA CACATAGATA TAGAGCAGA ANTANTAGAGA CACATAGATA TAGAACACA ANTANTAGA TAGAGCAGA ACACACAGA ACACACAGA ACACACAGA ACACACAGA ACACACAGA ANTANTAGATA TAGACACACA TAGACACACA TAGACACACA TAGACACACA TAGACACACA TAGACACACA TAGACACACA TAGACACACA TAGACACACACACACACACACACACACACACACACACACA	GT 18007 181
INGRAMATA CAGTAGCTTE ANTRACACA ATTANTITIC TAGATIGGAT CAIATITHA ATALACTAI GIAATCAT CACACCTOAL TICTATCA INCRIMITE TCCAMITITI TICTCACACT ITALACTACA ANAGCATIA ATTITIGGATA ITALTICAGA TATATICAGA TATATICAGA CACACCACACACACACACACACACACACACACACA	GT 18007 AT 18107 AT 18107 AT 18207 AT 18507 AT 18507 AT 18607 AT 18607 AT 1907 AT 2007 AT 200
INGALATAT CAGTACCTIC MATACACCA ATTANTITIC TAGATIGCAI CATATITIA ATATACATA GTANICATA ACACCIGA ITCITICE STORTINITY COCANTITY TICTCATATITY TICTARA TITATACA MAGCARTA ATTATACATA MATATACAC STORTANIA TATANICATA MATATACACA MATATATACA GALAGGAGTA TITCACATA GALAGGATA ATTATACATA MATATACAC STORTANIA TATANICATA MATATACACA MATATATACA GALAGGAGTA CACCTIGATA GALAGGATA TICCAGANACA TICCAGCATAC STORTANICATA MATATACACA MATATATACA GALAGGAGTACAC STORTANICA GALAGTATA TITTATITY GACCTICACAC STORTANICA GALAGTATA TITTATITY GACCTICACAC STORTANICA GALAGTATA TITTATITY GACAGTAGACA STORTANICA GALAGTATA TITTATITY GACAGTAGACA STORTANICA TITTATACACAC MACCACACA STORTACACAC STORTANICA TATANACACAC ACACCACACA ATALAGGATA GALAGTATAC STORTACTATI CAGTACCACTA TACCACCAC STORTACTATI TATANACACAC ACACCACAA ATALAGGAGTA GALAGTATAC STORTACTATI TATANACACAC ACACCACACA ATALAGGAGTA CACTAGAGGAGA STORTACATA TATANACACA CACCACACAA ATALAGAGACA CACTAGAGAGA STORTACATA TATANACACAC ACACCACACA ATALAGAGACA CACTAGAGAGAC STORTACATA TATANACACAC ACACCAGACA ATALAGAGACA CACTAGAGACA CACTAGAGACA STORTACACACACACA CACCACACAA ATALAGAGACA CACTAGAGACA CACTAGAGACA STORTACACACACACACA ATALAGAGACA ATALAGAGACA CACTAGACACACACACACACACACACACACACACACACAC	GT 18007 181
INGALITAT CAGTACCTIC ANTACACCA ATTACTTIC TAGATIGCA CALASTITIA ATALACTA GRANICATE ACAACCIGA ITCITICE STORTANTO TCCANTITT TICTCATACA TITATACCA ANGCARTA ATTACTGAN ATTACGANA ATTACTCACA TICATACCA STORTANTAC TATACCATA ATTACCATACA TITATACCA ANGCARTA ATTACTGANA GACGATIAN ATTACCACA STORTANTAC TATACCATACA ANTALATICAG TGALAGGAT CACATIGGA TICAGACAGA TICAGACAGA STORTANTAC TATACCATACA ANTALATICAG TGALAGGAT CACATIGGA TICAGACAGA TICAGACAGA STORTANTACATACATACACACACACACACACACACACACACAC	GT 18007 181
REGIMENTA CAGAGACTIC MATRACACA ATTANTITIC TAGATIGCAT CATATITIA ATATAACTA GTANICATA ACAGCIGA ITCITICE STORTANIA CATANICATAA ANTANICATAA ATAATICAA TAGACATTAA ATTIGGTGAA TITCECATA TIGATICATAA ATTAATICACA MATRATICAA TAGACATTAA ATTIGGTGAA TICCACAGA TICATACTAC GAGGATAAT TITTATICACAA ATAATICAG TAGATAGAT CAGACITAGA TACACTAGATAAT CAGACATATAA AATAATICAG TAGATAGAT CAGACTAGA ATTICACACA CAGACTATAA AATAATICAG CACACTAGATAA TACACTAGATAA TACACAGACA ATTICACACA CACACTAGATAA TACAGACATAA AATAATICATA TITCACATAA TAGACTAGATAA TACACTAGATAA TACACTAGATAAA TACACTAGATAA TACACTAGAA TACACTAGATAA TACACTAGATAA TACACTAGATAAAAAAAAAA	GT 18007 181

# Figure 8D

ACTTCAGCCT	GAGTGACAGA	GTAAGACCCT	ATCTCAAAAA	ACAGAAAAA	AAAAACACTG	GCCCAAAGGA	AATGAACTTG	TTACAGAAGC	CGGGGTTTAN	•••
AACACCAAAT	AATGCACTTG	TACCTAGTCC	TTCCCGGGTG	CTCTGCAGAC	ATTTCTCCAA	GCGTAGTCTG	CAAACAACCT	ACATATGTAG	AATTACCE	21:
ARTRACCOLC	CROMMOCCO	CCCCCCCC	CCC1 CC1 1 CC	G1 G1 G1		ACC THEWALT	WANCECTARG	ACATTCCGTC	TCTGAGARA	21e 217
ATTTALLTET	AAATTAACTA	CALTRADORC	A	-			CONTINUEIN	<b>NULLACTAGEE</b>	ACATATOO	216
CTACTGTAT	TGGAGAGTGC	AAGCGGAGAT TTAGTAGTAA	AGAACACTCT	ATTACTGCAG	AAATTTCTAT	TGGATAGCAC	TTATAATAGT	TTAGTGTAAC	TTAAAACTCC	215
TCARACTGGT	GACAATTTAA	CCCACAARCA	CCOLLOCOCO		WASTIVENI	CICIOCCUII	AUMUNCTUAT	WYLLLYWY	TTTGCAATTA	220 221
GAATCCCTGC	CATTTTTCAA	CATALTICAT	COLLCON	1010101010	*************		WE INDCIDIO	TATUTATUCA	GGATTCTTCA	222
TGGGACACTG	TATCHTCCTT	TTTLCCTCCT	CRCTMACALA			WANDER WATE	1110CIGGAC	CIGNICITAL	AACTCATAAA	223. 224
CATATCTTTA	CTCTTCTTCA	CAACTOCTOCTA	1-1			CIGCIICCCI	TACCTACCAT	WCCCCLLCC	TTCACTCATC	225.
ATGTATTTAA	TATCCATGTA	TCTATTCTCT	CTAATTTTGT	CATTITGTGT	TCTCATGTAT	TITCATTCAT	TATCTCTCCA	ACTOCATOR	TATGCCTTCA	226:
TACAACAAA	GATAACTTCC	TATGACAATT	ATCTTCCTTG	CCTTTCTCCC	ACATAGAACA	GTGCACAGAG	TAGGGGATCC	AAGAACCCAG	GAGAATATAT	227: 228:
										229:
17.01.0 311.0 0.0	2444704100	-under over 1 M	INGILLENGE	TALACGGGAG	SCIGAGGCAG	CICILIFICACE	BC LLCCCCCC	1000000000		230:
AAGGATGTCT	GCTTTGTGAG	CCAGCCTGGG	CCACAGAGCG	AGACTCCGTC	TCAMMANA	MAMAMA	AMAGTCCAAG	AAAAATTTTA	*****	2315
AAGGCTCAAT	TTAGTCACAT	TITAGCATIG CATTICCGTT CITATITICT	TCTCACCCAC	CCCCTTTAAA	TGAAATGGCA	AATACATTTA	AATCAGAACT	AAAAAGGGGA	ACAGGGTATA	232!
TTAGTAACTC	AAGCAGACAC	CTTATTTTCT	TTTCAAGCAG	AAAAGACTAT	GAGATGGTCG	TIGIGGITGE	TCCGGGAGGG	AGRAGATATA	GAAAGAAAGT	2335
GATCACTCAT	ACATGCATGT	GACCTCACTG	CACACTTATA	GTTATTGTAC	CTGTTGTCTT	TTTGCTGTCA	AGCCTAGCTA	AGATCATTTG	GAATGTTCAA	234! 2355
ACTITITATIC	TTTTCCAAAG	GCALCARCOT	C) CCC) CCC	WIGHT COUNTY	CCINITICAT	CCACATGAAC	TAAGATTACT	GATGTGTACA	GATTCAAAGC	3363
TCCAGTTCCT	TATGAATGGT	TACTGGTTTT GGATGGGCCA	CALLALATATG	AGATAAATTG	AGTGTATAAA	AGTCATTTT	AGACAAAATG	AAACAGGAAA	ACCACCICCA	2375
CAGAATCTCT	CCTCATTTGT	GGATGGGCCA AAGAGAAAGC	GCTCCACCAT	GTCATGGTTA	ATCTGCAGGG	AGGAAATACT	AGATTTCATT	GCAGATCAGA	CTGCAGCAAA	2385 23 <b>9</b> 5
CAGTGTCACC	TAGAAAAGAG	TOTTTCAAAA	WCCCC A CCC	0100000110	MATAGLANN	ACAITATATA	TCTAGCTTTG	AAATATGAAA	TACTGTTTAG	2405
AAATAAAGTG	ATCACTTGGT	GAAGAAATCT	CACAAAGAAG	AACATAGAGA	GTTCACTTTC	ATCTGGAGTA	ATGARCAGAT	TICANATING	CCAGGGTGGG	2415
GCAAGACGAT	AGAMAAGGTG	TAGGTGAGCT	GTTTGCAAGA	GCCACAAGGG	AAAGGGGAAG	ACAACTICTT	TGTGGACTTA	AGGGTGAAAG	TTGCAAGCAG	2425 2435
CARATGITTG	TOGGAATTGT	TGACTTAAAC	ACCECEMENT	CCUUCUNCEN	CIGGOTTOGT	TACGUAGGTT	GGGCAGCATT	GGGAGCAAAT	GTTGATTGAA	2445-
CCAGAGATCA	GAGCAGGCTA	AGGGACTGCT	GGGATCCTGT	CCAGCTTTGA	GACCCTACAG	AGCCATGTTC	ACCTAGGGA	GAGGGCATTT	GTTCACCTGG	2455
CCCCCTATCC	CCTTATTCCA	GGGCTTTCAC	CTCAGCTTGC	CAGGCTGGAG	CCAAGGGCCA	ACGCAGCCGC	CCCTTGTTCG	CGATGGTAGC	TTCCCAGGAG	24651 24751
GTTAGAAGGT	TCCGGACAGG	AACGGCGTGA	CCCCAATCCA	TIGGIAGETE	CINANGCCAN	AGGCACTGGC	GGGCCGGCC	AGCTTCTAAA	GTCGCCCAAG	2485
CGGGAAAGGA	AGCAGGGTCT	CTGAAGAAAT GTGTAGTAAA	ACTTCAGGAG	TAGAAAGAGG	AAGCTAGAGG	GTTAAATGCA	CAGCGATGGG	CTCAGAGCTC	CTTGAGAACT	2495;
AGGGAAACTG	ATGTCTAGAG	GTGTAGTAAA	CTAAAACAAG	TCTTGAATTG	CATACCGCCA	CGTAGGGAAG	AAATGAAAAC	CTTTGAATAT	TAGTGAAAA	25057 25157
AATTTGGTTT	GGATCCCATG	CCCATGACCC	TOCCACCACA	CT TENOR LO	CARCONCAGA	TTTAAAGAAG	CAACACCGCA	TITIGGCTTT	CTANAGCTIT	25257
TGGCCCTTTA	TGTGAAGTAC	CTGGTTTTTC	CATTTTCTGT	TTTACCATAG	GCCTCAGTTC	GGTGTGTGGC	CTATTTATTC	CACATTTCCG	AAGAACTATT	25357
TCATTCTATT	AGATTAAAAA	AAAAGAATAC AATGTCCAGG	AATGGAAGCC	AAGTGATTAA	GCTTTCCTTA	TGCTTATATT	AAGTTGTAGC	ATATGCATTT	ACCGATAGTE	25451 25557
TCTCCATCCA	CTTCCCTCAG	CTTTGGCCTC	AACCEAGCE	WATTO TIME	CIMITITION	CCTAAAGAAA	AICTITAAAA	TGTCTTAGCA	TTTTCCCCAG	25657
TGTTCGATTA	GGACACATCT	CAGTGGCAGA	TAACATCCAR	TOWN GOING	CIGINCINGC	TCTTGCCCTG	TACAGCTAGC	TACAGAGATT	CAATCCTTTC	25757
GTCTTAAGAC	TATAGTAATA	TCTTCACTTG	AAAAAGCCCT	CTATTATTCC	TATCTCAGAT	GATAAAAATT	CAATTAAGAG	AAATAAGAAC	GTGACATGTG	25857 25957
MAMACCAAA	GTGAGCATCC	CATCTGTTCC	CACTCALLTC	I TORRUMENT	ATCCTCTAAG	ATTATTEAT	CAAATGCATT	TCAATGACTA	GTTAACCATT	26057
CAAATGAATT	TECTTTETAT	ATGAGTGAGA	CCALACACTC	MACHINA MACHINA	VVGGVC LVGG	CAMACCACAT	CTGTGGGCAT	AGCAAGCTGT	ACATCACAAA	26157
										26257 26357
TTGACAGATT	ATAACTCAGA	TGTCTTACTC	AGAGCATATC	CCTTCCCATT	TCTGAAAGAG	GGAAAACAGG	CTCCCATTAG	ACTATGACTA	ACAAAAATGT	26457
GCAGACATCT	CATACCCCAA	ATAGCTAATA TEGCAAGGAT	TITTGATAGE	TATGATCCTG	AACGGCCAAA	CATTCCAAAA	CCAACTACTT	AAGAACTTGT	CCTTGACCGA	26557
CACCAGAAAG	GCCTTTTCCT	TGGCAAGGAT	GTTTGGTCAG	GGGTTGGCAA	AAATAATGCT	CTTCAGACTT	AAAAGAACAC	AACCATATTT	CTTAGCCATC	26657. 26757
CCTGGCCTAA	CTAGCCTACT	GAGCTGAGAG	ATCTCCAATC	TCCCCCCAR	CACTACCTAC	ATCATAATCT	CICIGCCCAG	GGGCTGTGGA	TGTCATCCAT	26857
									GAATTCTTAA	26957
AGCCAGACAC	ATGGTCTTAT	GACCGGCGTA	CTTACGCAGG	GCTTTGCACT	GAGACAGGTC	GTGCATCTGA	GGTTTACTGC	TITCCATTT	TGTTTTGTAA	27057 27157
TCCCACTCCC	CTACCACCTG	GTGCCTTAGC	CAGCCCTAGG	GACTTGAATC	CCTGCAGCCC	CATTTCACTT	CTCACCACCT	TCCGGGGTGG		27257
CCTCCTCATC	CCTACCTAGC	TACCATTGCC	ACTCCCCTCC	ECCAGCGGGG	ACATGGGCAT	AGGAGCAGGG	AGAGTTAAGG	TEGTECCEAG	ACATGCTCTC	27357 27457
										27557
AGATTGACAC	ATCATCCCTG	CTGGGGCCCCA	CTCTCTCTC	CCCCCCCCCC	TGGGAATCAC	CTCTCCACCT	TGATTGCCAC	AGTAGGCCAG	TGACAAGGGA	27657
										27757 27857
										27957
CAGGTCTCAA	ACTTAATTAT	AGCCATTCCT GTAAAGAATA ATCCCCAGAT	TTCTGGAGAG	AGGACTITTA	AAACATAAAG	ATTATCAAGT	CTTGGAAATT	CTGATTCAGT	AGATATATAA	28057
										28157 28257
										28357
GAAAATTGTT	AAGACTGGGT	TGTATGCACA	CTGCTGTTTA	TTATATATA	CCACAATGAT	TAGCCCAAAG	TARACACTCA	ATAAATGTTC	AAAAATTTAG	28457
		GIGCHGICGI	CUCOCCIGIA	ATTUCKAGUKE	TTTGGGAGGC	CAAGGCGGGC	GGATCACGAG	CTCCACACAC	CCLCACCASC	28557 28657
FIGULENACK	TEGTERARCE	CCATCTCTAC	TAMMATACA	AAAATTAACT	GGGCATGGTG	GCATGCGCCT	GTAGTCCCAG	GAGAATTCCT	TOLLCOTOC	28737
MONCOCK T	TOCAGTGAGE	CAAGATCTCA	CCACTGCTCT	CCAGCCTGGT	GACAGGGCAA	GACTCCGTCA	ALALABALA	AGAGAGGGAG	ACCCACACTA	28857
	AGTLAGAGCC	CTTTAATGAG	TOACCTTTCT	ACCTOTOCAL	COLCCICCO					28957
										29057
CCAGCTTTCT	TCCTATAAAG	GTGGATCAAG	GCACTTGCTT	ACAACTCCAA	CTGAAAAGGA	TTTACCAAAT	GTTGAGTGTG	CCCTCTAGTG	TTCACACTTC	29157
										29257 29357
										29457
MANGACCTC	CTAGAAAAAT	CAGTAGTTTT	TCTCTTTTCA	CHARGECETT	ATGGAGGCCT	TAGGICTIGC	TTCACAATAT	TCCAGTTTGA	AAAGGGTTTG	29557
										29657 29757
										29157
		AAAACACTAA	- I CHITCITECA	TADATTOUR	TTCCAGAAAC	ATTCCATTTC	TGCCAGCACC	TAGAAGCCAA 196	TATTTTGCCT	29957
ATTCCTGTAA	CCAGCACACA	TATTATTIT	TTTCTAGATC	AAATGTATTA	TGCAGTAAGA	GTCTTÄATTP	TGTTTTCACA		eu Ash Gly	30053
Lys Val Ass	Ala Phe Co	ra Glu Glu A	Per Tla Val	Ace (10 to				***		
		33	TCT ATC GTT	AAT GAA AA	N TGG ATT G	ta act set (	CC CAC TOT	CTT GAA ACT	ंदर्भ देश	30134
NA ATT AC	ं क्षेत्र केंट वे	CA GAGTANAT	ACACAGAAAG	AATAATAATC	TGCAGCACCA	CTAGCTCTTT	AATATGATTG	GTACACCATA	TITTACTAAG	30229

Figure 8E

GETANTANA ATTETTETTE ANTANATIEG GETANAGERA GANGGETCAT ANTITEAGAN CECNEGTEGE ACCEPTECTE ANGENTECHT AGTICITITE MINICECET ATTATEACTE ATTECHTE GETACANTA GETETCHTE TAGECATTITE CATACEAGAN GGECTTECEN AMARICAGTE TENTOTEAC GETENTIATA TETTETGGTGET TEGTECHANC CHORTOGEN ACCEPTATA TETTETGGTGET ATTACEAGE TEGTECHANCA GANGGEGT ACCEPTAGENTE ANTITECHTE ANAITTIGGAT ANTITEGGT ANAITTIGGAT ANTITEGGT ANTITEGGT ANAITTIGGAT ANTITEGGT TETTGGTCTG ANAITTIGGTC ATTACEATTA ACCAMANTA TEACANTATA AGANTGAGT CTITAACAT 234 GI) GIU	30429 30529 30529 30629 30729
CONTINGE TEAGRETICE CANGINGTEN CITAGAMANT CIGIGIATGE GANANCIGI ITGIGACITA AMATGAMANI INTITITANI AG GI GAN	30826
HIS AST ILE GIU GIU THE GIU HIS THE GIU GIN LYS AEG AST VAL ILE AEG ILE ILE PEO HIS HIS AST TYF AST ALA ALA CAF AAT ATT GAG GAG ACA GAA CAT ACA GAG CAA AAG CGA AAT GTG ATT CGA ATT ATT CCT CAC CAC AAC TAC AAT GCA GCT	30907
HE AST LYS TYT AST BIS AST HE ALS LOU LOU GLU LOU AST GLU PTO LOU VAL LOU AST SOT TYT VAL THE PTO HE CYS HT ANT ANG TAC AND CAT GAC ATT GCC CTT CTG GAA CTG GAC GAA CCC TTA GTG CTA AND AGC TAC GTT ACA CCT ATT TGC	30988
HE AND AND LYS GIU TYF THE AND ILE PHE LEU LYS PHE GIY BEF GIY TYF VAL BEF GIY TYP GIY ARG VAL PHE BIS LYS HT GCT GAC AAG GAA TAC ACG AAC ATC TTC CTC AAA TIT GGA TCT GGC TAT GTA AGT GGC TGG GGA AGA GTC TTC CAC AAA	31069
Cly Arg Ser Ala Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg Ser Thr Lys Phe Thr CCC AGA TCA GCT TTA GTT CTT CAG TAC CTT AGA GTT CCA CTT GTT GAC CGA GCC ACA TGT CTT CGA TCT ACA AAG TTC ACC	31150
lie Tyr Ann Ann Net Phe Cys Ale Gly Phe His Glu Gly Gly Arg Anp Ser Cys Gln Gly Anp Ser Gly Gly Pro His Val MC TAT AAC AAC ATG TTC TGT GCT GGC TTC CAT GAA GGA GGT AGA GAT TCA TGT CAA GGA GAT AGT GGG GGA CCC CAT GTT	31231
the Glu Val Glu Gly The Ser Phe Leu The Gly Ile Ile Ser Tep Gly Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile MT GAA GTG GAA GGG ACC AGT TTC TTA ACT GGA ATT ATT AGC TGG GGT GAA GAG TGT GCA ATG AAA GGC AAA TAT GGA ATA	31312
415  Tyr Thr Lys Val Ser Arg Tyr Val Asn Trp 11e Lys Glu Lys Thr Lys Leu Thr STOP  INT ACC ANG GTA TCC CGG TAT GTC ANC TGG ATT ANG GAA ANA ACA ANG CTC ACT TAA TGAARGATGG ATTTCCAAGG TTAATTCAT	_
GEARTIGARA ATTRACAGGG CCTCTCACTA ACTRATCACT TTCCCATCTT TTGTTAGATT TGARTATATA CATTCTATGA TCATTGCTTT TTCTCTTTA	C 31499
AGGGCAGAAT TICATATITI ACCIGAGCAA ATIGATIAGA AAAIGGAACC ACIAGAGGAA IAIAAIGIGI IAGGAAAITA CAGICATITC TAAGGGCCC	A 31599
COCTIGACA ANATIGIGAN GITANATICE CCACTCEGE CATCAGATAC TATGGITCEC CACTATGGCA ACTAACTCAC TCAATTITCC CTCCTTAGC	A 31699
CANTOCATO TOCCCGATOT TOTTTGCTTC TOCAACCAAA ACATCAATGT TTATTAGTTC TGTATACAGT ACAGGATCTT TGGTCTACTC TATCACAAG CLGTACCAC ACTCATGAAG AAAGAACACA GGAGTAGCTG AGAGGCTAAA ACTCATCAAA AACACTACTC CTTTTCCTCT ACCCTATTCC TCAATCTTT	G 31799
MINISTER RETURNS ANGENERY GONGLECT REAGGETANN RETURNERS AREACTRETE CHTTTCCRT RECETATION TEXASTER MINISTER RECETATION AREACTER TEXASTER TEXASTER FOR THE TEXASTER TEXASTER FOR THE TEXASTER TEXASTER FOR THE TEXASTER TO THE TE	T 31899
TICTOTTATA CHICTOTACA CAGTTATACA TICTOTACAA ACCCAGACTT GCTTCCATAG TGGAGACTTG CTTTTCAGAA CATAGGATG AAGTAAAGGA	A 31999 G 32099
CTGALAGT TTGGGGGAAA AGTITCTITC AGAGAGTTAA GTTATTTTAT ATATATATA TATATATA	G 32199
PROTECTION TOCOTECTION TAGACACACA COCATACACA CATATAATOG AAGCAATAAG CCATTCTAAG AGCTTGTATG GTTATGGAGG TCTGACTAC	G 32299
CHIGATITCA GGAAGGCAAG ATTGGCATAT CATTGTAACT AAAAAAGCTG ACATTGACCC AGACATATTG TACTCTTTCT AAAAATAATA ATAATAATG	C 32399
TAKAGAAAG AAGAGAACCG TTCGTTTGCA ATCTACAGCT AGTAGAGACT TTGAGGAAGA ATTCAACAGT GTGTCTTCAG CAGTGTTCAG AGCCAAGCL	A 32499
CHACTTONNG TTGCCTAGAC CAGAGGACAT AAGTATCATG TCTCCTTTAA CTAGCATACC CCGAAGTGGA GAAGGGTGCA GCAGGCTCAA AGGCATAAC	T 32599
CHICCANTC AGCCARCIAN GITGICCITT TOTGGITTCG IGTICACCAT GGARCATTIT CATTATAGIT ANTICITOTA TOTIGAATCI TOTAGAGAG	7 32699
PRIGACCAA CTGACGTATG TITCCCTTTG TGAATI <mark>AATA AA</mark> CTGGTGTT CTGGTTCÅTA CCTTGGCTTT TIGTGGATTC <u>CATTG</u> ATCTG AATCAGTC	C 32799
CHITATTY ATGATGCATG GGACTACTGA CAAAATCACT CTGACCCTGC CAAGCTGCTG CCCTCTCTCTG CCCCAACCTC ACCCCCAGCC AGGCCTCAC	T 32899
CHOCHAGTE COTTEAGTIC TITTAGECAA TATATITTING TOTTCGCATA TAAGTATAA TAAACATATT TITAAATTIC TINGGCTGGGC CONGTGGCT	
MOCTATAL TOCCAGOACT TOTGGAGGCC AAGGTGGGCG GATCACCTGA GGTTAGGAGT TTCAGGCCAG COTGGCCAAC ATGGTGAAAC COTGTCTCT	
TANAMATAG AACAATTAGC TGGGCTTGGT AATGTGCACC TATAATCCCA GCTACTGGGG AGGCTGAGGC AGGAGAATCA CTTGAGCCTG GGGAGCAGC	
STOCGGGGG GTTGCAGTGA GACAAGATCG CACCAGTGCA CTCCCCATCC TGGGTGACAG AGTGAGACTC TGTCTCAAAG AAAATAAATA AATAAATA	
THICTIGAGG COTTTCTTGT TARATCATTC ATGGAGAGGC ATCCCARACA CCACATTCAL CARACACTC TGARARATGT TTTCARATGC ARTATACC	C 33399
MELGAGATT TGATGCTCTG TTATCCAGTT TTCATATAAG GCTGTGTGAG CTGTGTCCCA GAGAGGACAG TGGTCTGAAT CCACCTGAGA CAGAATTGG	G 33499
MINACTAC TOTGAGTATG GOOTTCAATA AGTCACTOTC CATTTGGGAA TITGATTTCT CCACTTGTAT AATGAGAGTA TITGACAGGA TGCTCTCCC	:A 33599
MICCOTICC AATTITCITA CICTOTGATT TCATCITTTI ATTITATIC CITCATCCAA CAAATACTCA AGGACTAATI CCTGTGTGCC AAATACCAI MINITCATI AAATTOTAAT TCAGAITTTA TATATATATA AATAATGTAT AATGTGTATA AATTGCTITG TGAGTGCCTA CTACACTGCT AGACAGTAG	C 33699
RETCHTUR CHICATERS GRATCHGART CONTESTED CONSISTER CHARTRIGICS TECHNICAGE ALCOTORANG ANGOCONCA THANKANAMAN	A 33899
ATMITECTIC GEGITTAGGI TITAIGAAAA AAIGAAAGGA AAITAGTICT GCTITITGTIC ACAAAGGAA GGGAAGAGAG AAGAGACACI ATAAITGTI	T 33999
SCITCAGATT TAAGGAGGAG GCTAATTCAT GCATTAAACA CGTTACTTCA AATTTGAATG ACCAAAGGTC TGTAGCCTCA GCACTTCAAA ATTGGTAAL	LA 34099
STANGACACT CTGGCCTTGT TTCCATAGAG ACCACCCCTT ACAAAGGCAC CAATGGGAAA CTGGCCTCAG GACTCCTGTT ATTGGTCTTC TCTGTGGC	G 34199
MAAAGGAGC TCTTGGACC ATAAATCTCT GAGCAAGAGT TCTTTTTTGCC ATGGGCTCAA AAATGATTAA ATTCATCATG AGCCACCTGT GGCATATT	:C 34299
GRACTARAC ATGTGGGGCC TITAGGCTCA CTAAGAGCCA ATGTCTTCAG AGCCAGCCCT GGCTTGATTC TACCTAGGGC ATTTGCAGTT GCCATATA	NG 34399
UNCATTAGT GCTTTCAAAA TTACTGTAGA TACTTTGCCT AAATAGACTA AAACATGCTG CCGTCATATT GGAAGTGACA GATTAAAATA GAACTCTTT UAGTGAAGG AAAGTGTGCT AATATAATGC AGTCATTTTA ACTTGCTGTT TAAGTGTGAT TGTTTTTAGT TCTTTTGAAT ATTATTTGTT TTATACTG	C 34499 C 34599
MELICANG PACTOTICAL MILITATICAL ANTANTHIA ACHARAGOT ITCTTCCTA CITACOGA CCARACAGA CACAGOTTACA ANTANTICC	TA 34699
TAINIANTIG CTARACAAGT TOOGAATGCT TACAGTCTAA TOOAAGAATG TOAGAGCTGC AAGGCCCTT AAACACCATC CAATCCACTC CACTCATT	PA 34799
CALATELLG AGATTEAGGG CAACATAAGG CCAGGCCCAA GATAACACAA TGACAGCCAG GACTAGAGCT CAAGTCTCCC ACCCTGCACT TTGAAAGA	NT 34899
MINITICA ACTGGAGTAC ATTAACTCTA CTGTCTATAT TTTTAGGGCA GCTGGGGCAT TCTGCATTGG TGGCAATCCT CTCAACAACC CTGGGACTA	JA 34999
MACTECETE GAATTCTTAC TAACAATTCT CTAATTGACC AAAAGGTGAC GAAATCAAGG AGACCAATAA GGTAGCCTTG GAAAGCAAGA GTGGC	35094

# Figure 9A

1	gaattccgtg	gatgtgcttt	aaaaccacac	ctaacgtttg	agcacaagtc	tcacgaactg
61	gcgctacaac	ttcatcagaa	acgaagtctc	caaatctgtc	caacgcaaaa	acacaaagga
121	gtctaatgac	taagtcttcc	aaccacaact	gtctgctgcg	cccggaaaac	aagccggggc
181	tctggggacc	cggggctcag	gccgcctcgc	tccggcctag.	ccccgccacc	ttagttgtgt
241	cateceegg	gcatgctgag	catccccccg	cggctccggc	acagacgccc	ggacctcagg
301	tctctgcctc	cgcgcggggg	cccggccctg	tggccggagg	gageggeegg	atqqaqcqqa
361	ggatgaaagg	cggatacttg	gaccagcgag	tgccctacac	cttctgcagc	aaatctcccg
421	gaaatgggag	cttgggcgaa	gcgctgatgg	tcccgcaggg	aaagctcatg	gacccgggct
481	ccctgccgcc	ttccgactca	gaagatctct	tccaggacct	cagtcacttc	caagagacgt
541	ggctcgcaga	agctcaggta	ccggacagtg	atgagcagtt	tgttcctgat	ttccattcag
601	aaaacttagc	tttccatagc	cccaccacca	ggatcaagaa	ggaaccccag	agtccccgca
661	cagaccccgc	cctgtcctgc	agcaggaagc	caccactccc	ctaccaccat	ggagagcagt
721	gcctttactc	cagacaaatc	gccatcaagt	cccccgctcc	cggtgcccct	ggacagtcgc
781	ccctgcagcc	cttttccagg	gcagaacagc	agcagagcct	cctgagagcc	tccagctctt
841	cccagtccca	ccctggccac	gggtaccttg	gtgagcacag	ctccgtcttc	cagcagcccq
901	tggacatgtg	ccactccttc	acatctcctc	agggagggg	ccgggaacct	ctcccagccc
961	cctatcaaca	ccaactgtcg	gagccctgcc	caccctaccc	ccagcagaac	ttcaagcagg
1021	agtaccatga	cccctgtac	gaacaggctg	gccagcccgc	ttcaagccag	ggtggggtca
1081	gtgggcacag	gtacccaggg	gcgggggtgg	tgatcaaaca	ggagcgcaca	gacttcgcct
1141	acgactcaga	tgtccctgga	tgtgcatcaa	tgtacctcca	cccagagggc	ttctctggac
1201	cctctccagg	tgatggagtg	atgggttatg	gctatgaaaa	atcccttcga	ccattcccag
1261	atgatgtctg	cattgtccct	aaaaaatttg	aaggagacat	caagcaggaa	gggattggag
1321	ctttccggga	ggggccaccc	taccagcgcc	ggggtgcctt	acaactgtgg	cagtttctgg
1381	tggccctgct	ggatgaccca	acaaatgctc	atttcattgc	ttggacaggc	cggggaatgg
1441	agtttaaact	aattgaacct	gaagaggttg	ccaggctctg	gggtatccag	aagaaccggc
1501	cagccatgaa	ttatgacaag	ctgagccgct	cgctgcgata	ctattatgag	aaaggcatca
1561	tgcagaaggt	ggctggcgaa	cgctacgtgt	acaagtttgt	gtgcgagccg	gaggccctgt
1621	tctctctggc	cttcccagat	aatcaacgtc	cagctctgaa	ggctgagttt	gaccggccag
1681	tcagtgagga	ggacacagtc	cctttgtccc	acttggatga	gagtcctgcc	tacctcccag
1741	aactcactgg	ccccgctccg	cccttcggcc	acagaggtgg	atattcttac	taggcaccag
1801	tggcttcccc	ttgacatggt	ggggttgctc	agtgtatata	tcaactgatt	tggtattggt
1861	gaaggccctc	tttctgatgc	ctgtagaagt	ctctggggtc	agagctccac	tatcccatct
1921	gatactcctg	gccagactca	gctgctaacc	agagtctgcg	ggaaagacag	tggaggcagg
1981	ccaaatctaa	aggcagtagc	tgaagttcgc	tgtggctcac	ctgtaccttc	agttcagctt
2041	ggcctctgcc	taggtcttgc	tcagaggcca	agttcctcac	ccccaccaca	gagatccagt
2101	gttctattct	ggggacatac	agggacttcc	cttgtttatt	atggcaacag	ggccaagggg
2161	attctcagaa	caccctgtgt	ctcccctctc	ccaacccccc	atgggagaca	aagttctgcc
2221	tggcttctgc	cctgaacagg	ggggtcctgt	gttcttggtg	ctgtgctctg	ggaggcagga
2281	gcatgtgggc	ggcagctggg	ggggggtgtg	gaagtagaga	tggctctctg	ccctaggcct
2341	acccaggcct	aattccacct	ttgcctctta	tgccagacct	taataaagcc	tctgcttctc
2401	cccggaattc					

### Figure 9B

MTKSSNHNCLLRPENKPGLWGPGAQAASLRPSPATLVVSSPGHA
EHPPAAPAQTPGPQVSASARGPGPVAGGSGRMERRMKGGYLDQRVPYTFCSKSPGNGS
LGEALMVPQGKLMDPGSLPPSDSEDLFQDLSHFQETWLAEAQVPDSDEQFVPDFHSEN
LAFHSPTTRIKKEPQSPRTDPALSCSRKPPLPYHHGEQCLYSRQIAIKSPAPGAPGQS
PLQPFSRAEQQQSLLRASSSSQSHPGHGYLGEHSSVFQQPVDMCHSFTSPQGGGREPL
PAPYQHQLSEPCPPYPQQNFKQEYHDPLYEQAGQPASSQGGVSGHRYPGAGVVIKQER
TDFAYDSDVPGCASMYLHPEGFSGPSPGDGVMGYGYEKSLRPFPDDVCIVPKKFEGDI
KQEGIGAFREGPPYQRRGALQLWQFLVALLDDPTNAHFIAWTGRGMEFKLIEPEEVAR
LWGIQKNRPAMNYDKLSRSLRYYYEKGIMQKVAGERYVYKFVCEPEALFSLAFPDNQR
PALKAEFDRPVSEEDTVPLSHLDESPAYLPELTGPAPPFGHRGGYSY

## Figure 10 (A)

1 gaattccagg ttggaggggc ggcaacetec tgccageett caggecacte teetgtgeet 61 gccagaagag acagagcttg aggagagctt gaggagagca ggaaaggtgg aacattgctg 121 ctgctgctca ctcagttcca caggtgggag gaacagcagg gcttagagtg ggggtcattg 181 tgcagatggg aaaacaaagg cccagagagg ggaagaaatg cctaggagct accgagggca 241 ggcgacctca accacagcc agtgctggag ctgtgagtgg atgtagagca gcggaatatc 301 cattcagcca gctcagggga aggacagggg ccctgaagcc aggggatgga gctgcaggga 361 agggagetea gagagaaggg gaggggagte tgageteagt tteeegetge etgaaaggag 421 ggtggtacct actcccttca cagggtaact gaatgagaga ctgcctggag gaaagctctt 481 caagtgtggc ccaccccacc ccagtgacac cagccctga cacgggggag ggagggcagc 541 atcaggaggg gctttctggg cacacccagt acccgtctct gagctttcct tgaactgttg 601 cattttaatc ctcacagcag ctcaacaagg tacataccgt caccatcccc attttacaga 661 tagggaaatt gaggctcgga gcggttaaac aactcacctg aggcctcaca gccagtaagt 721 gggttccctg gtctgaatgt gtgtgctgga ggatcctgtg ggtcactcgc ctggtagagc 781 cccaaggtgg aggcataaat gggactggtg aatgacagaa ggggcaaaaa tgcactcatc 841 cattcactct gcaagtatct acggcacgta cgccagctcc caagcaggtt tgcgggttgc 901 acagcggagc gatgcaatct gatttaggct tttaaaggat tgcaatcaag tgggacccac 961 tagecteaac ectgtacete ecetecete caceceage agtetecaaa ggeetecaac 1021 aaccccagag tgggggccat gtatccaaag aaactccaag ctgtatacgg atcacactgg 1081 ttttccagga gcaaaaacag aaacagcctg aggctggtca aaattgaacc tcctcctgct 1141 ctgagcagcc tagggggcag actaagcaga gggctgtgca gacccacata aagagcctac 1201 tgtgtgccag gcacttcacc cgaggcactt cacaagcatg cttgggaatg aaacttccaa 1261 ctctttggga tgcaggtgaa acagttcctg gttcagagag gtgaagcggc ctgcctgagg 1321 cagcacagct cttctttaca gatgtgcttc cccacctcta ccctgtctca cggccccca 1381 tgccagcctg acggttgtgt ctgcctcagt catgctccat ttttccatcg ggaccatcaa 1441 gagggtgttt gtgtctaagg ctgactgggt aactttggat gagcggtctc tccgctccga 1501 gcctgtttcc tcatctgtca aacgggctct aacccactct gatctcccag ggcggcagta 1561 agtetteage ateaggeatt ttggggtgae teagtaaatg gtagatettg etaceagtgg 1621 aacageeact aaggattetg eagtgagage agagggeeag etaagtggta eteteeeaga 1681 gactgtctga ctcacgccac ccctccacc ttggacacag gacgctgtgg tttctgagcc 1741 aggtacaatg actcetttcg gtaagtgcag tggaagctgt acactgccca ggcaaagcgt 1801 ccgggcagcg taggcggcg actcagatcc cagccagtgg acttagcccc tgtttgctcc 1861 tccgataact ggggtgacct tggttaatat tcaccagcag cctcccccgt tgcccctctg 1921 gatccactgc ttaaatacgg acgaggacag ggccctgtct cctcagcttc aggcaccacc 1981 actgacctgg gacagtgaat cgtaagtatg cctttcactg cgaggggttc tggagaggct 2041 tccgagctcc ccatggccca ggcaggcagc aggtctgggg caggaggggg gttgtggagt 2101 gggtatccgc ctgctgaggt gcagggcaga tggagaggct gcagctgagc tcctattttc 2161 ataataacag cagccatgag ggttgtgtcc tgtttcccag tcctgcccgg tccccctcg 2221 gtacctcctg gtggatacac tggttcctgt aagcagaagt ggatgagggt gtctaggtct 2281 gcagtcctgg caccccagga tgggggacac cagccaagat acagcaacag caacaaagcg 2341 cagccatttc tttctgtttg cacagctcct ctgtctgtcg ggggctcctg tctgttgtct 2401 cctataagcc tcaccacctc tcctactgct tgggcatgca tctttctccc cttctataga 2461 tgaggaggtt aaggttcaga gagggtggg gaggaacgcc ggctcacatt ctccatcccc 2521 tccagatatg accaggaaca gacctgtgcc agcctcagcc ttacatcaaa atgggcctcc 2581 ccatgcaccg tggacctctg ggcctcctg tcccagtgga ggacaggaag ctgtgaggg 2641 cactgtcacc cagggctcaa gctggcattc ctgaataatc gctctgcacc aggccacggc 2701 taageteagt gegtgattaa geeteataae eeteeaagge agttaetagt gtgatteeca 2761 ttttacagat gaggaagatg gggacagaga ggtgaataac tggccccaaa tcacacacca 2821 tccataattc gggctcaggc acctggctcc agtccccaaa ctcttgaacc tggccctagt 2881 gtcactgttt ctcttgggtc tcaggcgctg gatggggaac aggaaacctg ggctgaactt 2941 gaggeetete tgatgetegg tgaetteaga cagttgetea acetetetgt tetettggge 3001 aaaacatgat aacctttgac ttctgtcccc tcccctcacc ccacccgacc ttgatctctg 3061 aagtgttgga aggatttaat ttttcctgca ctgagttttg gagacaggtc aaaaagatga 3121 ccaaggccaa ggtggccagt ttcctataga acgcctctaa aagacctgca gcaatagcag 3181 caagaactgg tattetegag aacttgetge geageaggea ettettggea ttttatgtgt 3241 atttaatttc acaatagete tatgacaaag tecaeettte teatetecag gaaactgagg 3301 ttcagagagg ttaagtaact tgtccaaggt cacacagcta atagcaagtt gacgtggagc 3361 aatctggcct cagagccttt aattttagcc acagactgat gctcccctct tcatttagcc 3421 aggctgcctc tgaagttttc tgattcaaga cttctggctt cagctttgta cacagagatg

### Figure 10 (B)

3481 attcaatgtc aggttttgga gcgaaatctg tttaatccca gacaaaacat ttaggattac 3541 atctcagttt tgtaagcaag tagctctgtg atttttagtg agttatttaa tgctctttgg 3601 ggctcaattt ttctatctat aaaatagggc taataatttg caccttatag ggtaagcttt 3661 gaggacagat tagatgatac ggtgcctgta aaacaccagg tgttagtaag tgtggcaatg 3721 atggtgacgc tgaggctgtg tttgcttagc atagggttag gcagctggca ggcagtaaac 3781 agttggataa tttaatggaa aatttgccaa actcagatgc tgttcactgc tgagcaggag 3841 ccccttcctg ctgaaatggt cctggggagt gcagcagget ctccgggaag aaatctacca 3901 teteteggge aggageteaa eetgtgtgea ggtacaggga gggetteete aeetggtgee 3961 cactcatgca tracgtcagt tattcctcat coctgtccaa aggattcttt tctccattgt 4021 acagetatga agetagtget caaagaagtg aagteattta eeccaggeee eetgecagta 4081 agtgacaggg cetggtcaca ettgggttta tttattgece agttcaacag gttgtttgac 4141 cataggegag attetettee etgeaceetg eegggttget ettggteeet tattttatge 4201 tectgggtag aaatggtgeg agattaggea gggagtggae getteeetgt eeetggeeee 4261 gcaaagagtg ctcccacctg ccccgatccc agaaatgtca ccatgaagcc ttcattcttt 4321 tggtttaaag cttggcctca gtgtccgtac accatggggt ccttggccag atggcgactt 4381 tetectetee agtegeeete eeaggeaeta gettttagga gtgeagggtg etgeetetga 4441 tagaagggcc aggagagac aggttttgga gacctgatgt tataaggaac agcttgggag 4501 gcataatgaa cccaacatga tgcttgagac caatgtcaca gcccaattct gacattcatc 4561 atctgagate tgaggacaca getgteteag tteatgatet gagtgetggg aaagecaaga 4621 cttgttccag ctttgtcact gacttgctgt atagcctcaa caaggccctg accetetetg 4681 ggcttcaaac tcttcactgt gaaaggagga aaccagagta ggtgatgtga caccaggaaa 4741 gatggatggg tgtgggggaa tgtgctcctc ccagctgtca ccccctcgcc accctccctg 4801 caccagecte tecaceteet tigageceag aatteceetg tetaggaggg cacetgtete 4861 gtgcctagcc atgggaattc tccatctgtt ttgctacatt gaacccagat gccattctaa 4921 ccaagaatee tggetgggtg caggggetet egeetgtaae eecageaett tgggaggeea 4981 aggcaggegg atcaagaggt caggagttca agacetgeet ggccaacacg gtgaaacete 5041 agetetacta aaaatacaaa aattagecag gegtggtgge acaegeetgt aateceaget 5101 atttgggaag ctgagacaga agaatttctt gaacccggga ggtggaggtt tcagtgagcc 5161 gagatcacgc cactgcactc caccetggcg gataaagcga gactetgtet caaaaaaaac 5221 ccaaaaacct atgttagtgt acagagggc ccagtgaagt etteteccag ecceaetttg 5281 cacaactggg gagagtgagg ecceaggace agaggattet tgetaaagge caagtggata 5341 gtgatggeee tgecaggeta gaagecaaaa eetetggeee tgagggeaatt eagaatatt 5401 agtgtcccca ccctgcagag gcccaactcc ctcctgacca ctgagccctg taatgatggg 5461 ggaatttcca taagccatga aggactgcac aaagttcagt tgggagtgaa agagaaatta 5521 aagggagatg gaaatataca gcactaattt tagcaccgtc ttcagttcta acaacactag 5581 ctagctgaag aaaatacaaa catgtattat gtaatgtgtg gtctgttcca tttggattac 5641 ttagaggcac gagggccaag gagaaaggtg gtggagagaa accagctttg cacttcattt 5701 gttgctttat tggaaggaaa cttttaaaag tccaaggggg ttgaagaatc tcaatatttg 5761 ttatttccag cttttttct ccagtttttc atttcccaaa ttcaaggaca cctttttctt 5821 tgtattttgt taagatgatg gttttggttt tgtgactagt agttaacaat gtggctgccg 5881 ggcatattet ceteagetag gaceteagtt ticecateig tigaagaegge aggitetace 5941 tagggggctg caggcaggtg gtccgaagcc tgggcatatc tggagtagaa ggatcactgt 6001 ggggcagggc aggttctgtg ttgctgtgga tgacgttgac tttgaccatt gctcggcaga 6061 geetgetete getggtteag ceacaggeee caccactece tattgtetea geecegggta 6121 tgaaacatgt attectcact ggcctatcac ctgaagcctt tgaatttgca acacetgcca 6181 acceptect caaaagagtt geeeteteta gateettttg atgtaaggtt tggtgttgag 6241 acttatttca ctaaattctc atacataaac atcactttat gtatgaggca aaatgaggac 6301 cagggagatg aatgacttgt cetggeteat acacetggaa agtgacagag teagattaga 6361 teetaggtet atetgaagtt aaaagaggtg tetttteact teecacetee teeatetact 6421 ttaaagcagc acaaacccct getttcaagg agagatgagc gtctctaaag cccctgacag 6481 caagageeea gaaetgggae accattagtg acceagaegg caggtaaget gaetgeagga 6541 gcatcagect attettgtgt etgggaccae agageattgt ggggacagee eegtetettg 6601 ggaaaaaac cctaagggct gaggatcctt gtgagtgttg ggtgggaaca gctcccagga 6661 ggtttaatca cagcccctcc atgctctcta gctgttgcca ttgtgcaaga tgcatttccc 6721 ttctgtgcag cagtttccct ggccactaaa tagtgggatt agatagaagc cctccaaggg 6781 ctccagcttg acatgattct tgattctgat ctgacccgat tctgataatc gtgggcaggc 6841 ccattcctct tcttgtgcct cattttcttc ttttgtaaaa caatggctgt accatttgca 6901 tettagggte attgeagatg aaagtgttge tgteeagage etgggtgeag gacetagatg 6961 taggattetg gttetgetae tteeteagtg acattgaata getgaeetaa tetetetgge 7021 tttggtttct tcatctgtaa aagaaggata ttagcattag cacctcacgg gattgttaca

#### PCT/US00/15728

### Figure 10 (C)

7081 agaaagcaat gaattaacac atgtgagcac ggagaacagt gcttggcata tggtaagcac 7141 tacgtacatt ttgctattct tctgattctt tcagtgttac tgatgtcggc aagtacttgg 7201 cacaggetgg tttaataate cetaggeact tteacgtggt gteaateeet gateactggg 7261 agteateatg tgccttgact cgggcctggc cccccatet ctgtcttgca ggacaatgcc 7321 gtettetgte tegtggggea teeteetget ggeaggeetg tgetgeetgg teeetgtete 7381 cctggctgag gatccccagg gagatgctgc ccagaagaca gatacatccc accatgatca 7441 ggatcaccca acettcaaca agatcacccc caacetgget gagttegeet teageetata 7501 cegecagetg geacaceagt ceaacageac caatatette ttetececag tgageatege 7561 tacageettt geaatgetet eeetggggae caaggetgae aeteaegatg aaateetgga 7621 gggcctgaat ttcaacctca cggagattcc ggaggctcag atccatgaag gcttccagga 7681 actoctocgt accotcaaco agocagacag coagotocag otgacoacog goaatggoot 7741 gttoctoago gagggootga agotagtgga taagtttttg gaggatgtta aaaagttgta 7801 ccactcagaa gccttcactg tcaacttcgg ggacaccgaa gaggccaaga aacagatcaa 7861 cgattacgtg gagaagggta ctcaagggaa aattgtggat ttggtcaagg agcttgacag 7921 agacacagtt tttgctctgg tgaattacat cttctttaaa ggtaaggttg ctcaaccagc 7981 ctgagctgtt teccatagaa acaagcaaaa atatttetea aaccateagt tettgaacte 8041 teettegeaa tgeattatgg gecatageaa tgetttteag egtggattet teagitttet 8101 acacacaac actaaaatgt tttccatcat tgagtaattt gaggaaataa tagattaaac 8161 tgtcaaaact actgacgctc tgcagaactt ttcagagcct ttaatgtcct tgtgtatact 8221 gtatatgtag aatatataat gcttagaact atagaacaaa ttgtaataca ctgcataaag 8281 ggatagtttc atggaacata ctttacacga ctctagtgtc ccagaatcag tatcagtttt 8341 gcaatctgaa agacctgggt tcaaatcctg cctctaacac aattagcttt tgacaaaaac 8401 aatgcattet acctetttga ggtgetaatt teteatetta geatggacaa aataceatte 8461 ttgctgtcag gttttttag gattaaacaa atgacaaaga ctgtggggat ggtgtgtggc 8521 atacagcagg tgatggactc ttctgtatct caggctgcct tcctgcccct gaggggttaa 8581 aatgccaggg tcctgggggc cccagggcat tctaagccag ctcccactgt cccaggaaaa 8641 cagcataggg gaggggaggt gggaggcaag gccaggggct gcttcctcca ctctgaggct 8701 cccttgctct tgaggcaaag gagggcagtg gaggcaagcc aggctgcagt cagcacagct 8761 aaagteetgg etetgetgtg geettagtgg gggeecaggt eeeteteeag eeecagtete 8821 eteettetgt eeaatgagaa agetgggate aggggteeet gaggeecetg tecaetetge 8881 atgectegat ggtgaagete tgttggtatg geagagggga ggetgeteag geatetgeat 8941 ttcccctgcc aatctagagg atgaggaaag ctctcaggaa tagtaagcag aatgtttgcc 9001 etggatgaat aactgagetg ccaattaaca aggggcaggg agcettagac agaaggtace 9061 aaatatgcct gatgctccaa cattttattt gtaatatcca agacaccctc aaataaacat 9121 atgattccaa taaaaatgca cagccacgat ggcatctctt agcctgacat cgccacgatg 9181 tagaaattot goatottoot otagtttiga attatoocca cacaatottt tioggoagot 9241 tggatggtca gtttcagcac cttttacaga tgatgaagct gagcctcgag ggatgtgtgt 9301 cgtcaagggg gctcagggct tctcagggag gggactcatg gtttcttatt ctgctacact 9361 cttccaaacc ttcactcacc cctggtgatg cccaccttcc cctctctca ggcaaatggg 9421 agagaccett tgaagtcaag gacaccgagg aagaggaett ccacgtggac caggtgacca 9481 ccgtgaaggt gcctatgatg aagcgtttag gcatgtttaa catccagcac tgtaagaagc 9541 tgtccagctg ggtgctgctg atgaaatacc tgggcaatgc caccgccatc ttcttcctgc 9601 ctgatgaggg gaaactacag cacctggtaa atgaactcac ccacgatatc atcaccaagt 9661 tectggaaaa tgaagacaga aggtgattee ccaacetgag ggtgaccaag aagetgecca 9721 cacctettag ceatgttggg actgaggee ateaggaetg geeagaggge tgaggagggt 9781 gaacceaca tecetgggte actgetacte tgtataaact tggetteeag aatgaggeea 9841 ceactgagtt caggeagege egteeatget ceatgaggag aacagtacee agggtgagga 9901 ggtaaaggte tegteeetgg gaactteeca etceagtgtg gacactgtee etteecaata 9961 tecagtgeec aaggeaggga cageageace accaeaegtt etggeagaac caaaaaggaa 10021 cagatgggct tcctggcaaa ggcagcagtg gagtgtggag ttcaagggta gaatgtccct 10081 ggggggacgg gggaagagcc tgtgtggcaa ggcccagaaa agcaaggttc ggaattggaa 10141 cagccaggcc atgttcgcag aaggcttgcg tttctctgtc actttatcgg tgctgttaga 10201 ttgggtgtcc tgtagtaagt gatacttaaa catgagccac acattagtgt atgtgtgtgc 10261 attogtgatt atgeceatge cetgetgate tagttegttt tgtacaetgt aaaaceaaga 10321 tgaaaataca aaaggtgtcg ggttcataat aggaatcgag gctggaattt ctctgttcca 10381 tgccagcacc tcctgaggtc tctgctccag gggttgagaa agaacaaaga ggctgagagg 10441 gtaacggatc agagagccca gagccagctg ccgctcacac cagaccctgc tcagggtggc 10561 atccactaaa cggttgtcac tgggcactgc caccagcccc gtgtttctct gggtgtaggg 10621 ccctggggat gttacaggct gggggccagg tgacccaaca ctacagggca agatgagaca

# Figure 10 (D)

10681	ggcttccagg	acacctagaa	tatcagagga	ggtggcattt	caagcttttg	tgattcattc
10741	gatgttaaca	ttctttgact	caatgtagaa	gagctaaaag	tagaacaaac	caaaqccqaq
10801	ttcccatctt	agtgtgggtg	gaggacacag	gagtaagtgg	cagaaataat	cagaaaagaa
10861	aacacttgca	ctgtggtggg	tcccagaaga	acaagaggaa	tgctgtgcca	tgccttgaat
10921	ttcttttctg	cacgacaggt	ctgccagctt	acatttaccc	aaactgtcca	ttactqqaac
10981	ctatgatctg	aagagcgtcc	tgggtcaact	gggcatcact	aaggtcttca	gcaatggggc
11041	tgacctctcc	ggggtcacag	aggaggcacc	cctgaagctc	tccaaggtga	gatcaccctg
11101	acgaccttgt	tgcaccatgg	tatctgtagg	gaagaatgtg	tgggggctgc	agcactgtcc
11161	tgaggctgag	gaaggggccg	agggaaacaa	atgaagaccc	aggctgagct	cctgaagatg
11221	cccgtgattc	actgacacgg	gacggtgggc	aaacagcaaa	gccaggcagg	ggctgctgtg
11281	cagctggcac	tttcggggcc	tcccttgagg	ttgtgtcact	gaccctgaat	ttcaactttq
11341	cccaagacct	tctagacatt	gggccttgat	ttatccatac	tgacacagaa	aggtttgggc
11401	taagttgttt	caaaggaatt	tctgactcct	tcgatctgtg	agatttggtg	tctgaattaa
11461	tgaatgattt	cagctaaagt	gacacttatt	ttggaaaact	aaaggcgacc	aatgaacaac
11521	ctgcagttcc	atgaatggct	gcattatctt	ggggtctggg	cactgtgaag	gtcactgcca
11581	gggtccgtgt	cctcaaggag	cttcaagccg	tgtactagaa	aggagagagc	cctqqaqqca
11641	gacgtggagt	gacgatgctc	ttccctgttc	tgagttgtgg	gtgcacctga	gcagggggag
11701	aggcgcttgt	caggaagatg	gacagagggg	agccagcccc	atcagccaaa	gccttgagga
11761	ggagcaaggc	ctatgtgaca	gggaggaga	ggatgtgcag	ggccagggcc	gtccaqqqqq
11821	agtgagcgct	tcctgggagg	tgtccacgtg	agccttgctc	gaggcctggg	atcagcctta
11881	caacgtgtct	ctgcttctct	cccctccagg	ccgtgcataa	ggctgtgctg	accatcgacg
11941	agaaagggac	tgaagctgct	ggggccatgt	ttttagaggc.	catacccatg	tctatccccc
12001	ccgaggtcaa	gttcaacaaa	ccctttgtct	tcttaatgat	tgaacaaaat	accaagtctc
12061	ccctcttcat	gggaaaagtg	gtgaatccca	cccaaaaata	actgcctctc	gctcctcaac
12121	ccctcccctc	catccctggc	cccctccctg	gatgacatta	aagaagggtt	gagctggtcc
12181	ctgcctgcat	gtgatctgta	aatccctggg	atgttttctc	tg	<b></b>

# Figure 11

1	caccagcatc	atctcctcca	attcatccag	ctactctgcc	catgaagata	atagttttca
<b>6</b> T	ggcggattgc	ctcagatcac	actatctcca	cttgcccagc	cctqtqqaaq	attaggggcc
121	atgtattcca	atgtgatagg	aactgtaacc	tctggaaaaa	qqaaqqttta	tcttttatcc
TRI	ttgctgctca	ttggcttctg	ggactgcgtg	acctgtcacg	qqaqccctqt	ggacatctgc
241	acagccaagc	cgcgggacat	tcccatgaat	cccatgtgca	tttaccqctc	cccqqaqaaq
301	aaggcaactg	aggatgaggg	ctcagaacag	aagatcccgg	aggccaccaa	ccaacatatc
36 T	tgggaactgt	ccaaggccaa	ttcccgcttt	gctaccactt	tctatcagca	cctqqcaqat
421	tccaagaatg	acaatgataa	cattttcctg	tcacccctga	gtatctccac	gacttttact
481	atgaccaagc	tgggtgcctg	taatgacacc	ctccaqcaac	tgatggaggt	atttaagttt
541	gacaccatat	ctgagaaaac	atctgatcag	atccacttct	tctttqccaa	actgaactgc
POT	cgactctatc	gaaaagccaa	caaatcctcc	aagttagtat	caqccaatcq	cctttttgga
661	gacaaatccc	ttaccttcaa	tgagacctac	caggacatca	gtgagttggt	atatggagcc
721	aagctccagc	ccctggactt	caaggaaaat	gcagagcaat	ccagagegge	catcaacaaa
781	tgggtgtcca	ataagaccga	aggccgaatc	accgatgtca	ttccctcqqa	agccatcaat
841	gagctcactg	ttctggtgct	ggttaacacc	atttacttca	agggcctgtg	gaagtcaaag
901	ttcagccctg	agaacacaag	gaaggaactg	ttctacaagg	ctgatggaga	atcatattca
961	gcatctatga	tgtaccagga	aggcaagttc	cgttatcggc	gcgtggctga	aggcacccag
1021	gtgcttgagt	tgcccttcaa	aggtgatgac	atcaccatgg	tcctcatctt	gcccaagcct
TOST	gagaagagcc	tggccaaggt	ggagaaggaa	ctcaccccag	aggtgctgca	ggagtggctg
1141	gatgaattgg	aggagatgat	gctggtggtt	cacatgcccc	gcttccqcat	tgaggacggc
1201	ttcagtttga	aggagcagct	gcaagacatg	ggccttqtcq	atctqttcaq	ccctgaaaag
1261	tccaaactcc	caggtattgt	tgcagaaggc	cgagatgacc	tctatqtctc	agatgcattc
1321	cataaggcat	ttcttgaggt	aaatgaagaa	ggcaqtqaaq	cagetgeaag	taccactatt
T38T	grgarrgerg	gccgttcgct	aaaccccaac	agggtgactt	tcaaggccaa	caggcccttc
<b>1441</b>	ctggttttta	taagagaagt	tcctctgaac	actattatct	tcatgggcag	agtagccaac
1501	ccttgtgtta	agtaaaatgt	tcttattctt	tgcacctctt	cctatttttq	gtttgtgaac
1561	agaagtaaaa	ataaatacaa	actacttcca	tctcacatt	J	5 5-5

# Figure 12 A

1	ctgcaggggg	999999999	gggggctgtc	atggcggcag	gacggcgaac	ttgcagtatc
61	tccacgaccc	gcccctacag	gtgccagtgc	ctccagaatg	tggcagctca	caagcctcct
121	gctgttcgtg	gccacctggg	gaatttccgg	cacaccagct	cctcttgact	cagtgttctc
181	cagcagcgag	cgtgcccacc	aggtgctgcg	gatccgcaaa	cgtgccaact	ccttcctgga
241	ggagctccgt	cacagcagcc	tggagcggga	gtgcatagag	gagatctgtg	acttcgagga
301	ggccaaggaa	attttccaaa	atgtggatga	cacactggcc	ttctggtcca	agcacgtcga
361	cggtgaccag	tgcttggtct	tgcccttgga	gcacccgtgc	gccagcctgt	gctgcgggca
421	cggcacgtgc	atcgacggca	tcggcagctt	cagctgcgac	tgccgcagcg	gctgggaggg
				caattgctcg		
541	gcattactgc	ctagaggagg	tgggctggcg	gcgctgtagc	tgtgcgcctg	gctacaagct
601	gggggacgac	ctcctgcagt	gtcaccccgc	agtgaagttc	ccttgtggga	ggccctggaa
661	gcggatggag	aagaagcgca	gtcacctgaa	acgagacaca	gaagaccaag	aagaccaagt
721	agatccgcgg	ctcattgatg	ggaagatgac	caggcgggga	gacagcccct	ggcaggtggt
781	cctgctggac	tcaaagaaga	agctggcctg	cggggcagtg	ctcatccacc	cctcctgggt
841	gctgacagcg	gcccactgca	tggatgagtc	caagaagctc	cttgtcaggc	ttggagagta
901	tgacctgcgg	cgctgggaga	agtgggagct	ggacctggac	atcaaggagg	tcttcgtcca
961	ccccaactac	agcaagagca	ccaccgacaa	tgacatcgca	ctgctgcacc	tggcccagcc
1021	cgccaccctc	tcgcagacca	tagtgcccat	ctgcctcccg	gacagcggcc	ttgcagagcg
1081						acagcagccg
1141	agagaaggag	gccaagagaa	accgcacctt	cgtcctcaac	ttcatcaaga	ttcccgtggt
1201	cccgcacaat	gagtgcagcg	aggtcatgag	caacatggtg	tctgagaaca	tgctgtgtgc
1261	gggcatcctc	ggggaccggc	aggatgcctg	cgagggcgac	agtggggggc	ccatggtcgc
1321	ctccttccac	ggcacctggt	tcctggtggg	cctggtgagc	tggggtgagg	gctgtgggct
1381	ccttcacaac	tacggcgttt	acaccaaagt	cagccgctac	ctcgactgga	tccatgggca
1441	catcagagac	aaggaagccc	cccagaagag	ctgggcacct	tagcgaccct	ccctgcaggg
1501						cacaccggcc
1561	tgctgttctg	tccttccatc	cctcttttgg	gctcttctgg	agggaagtaa	catttactga
1621	gcacctgttg	tatgtcacat	gccttatgaa	tagaatctta	actcctagag	caactctgtg
1681	gggtggggag	gagcagatcc	aagttttgcg	gggtctaaag	ctgtgtgtgt	tgagggggat
1741	. actctgttta	tgaaaaagaa	taaaaaacac	aaccacgaaa	aaaaaaaaa	aaaaaaaaa
1801	. aaaaaaaaa	aaaaaaaccc	ccccccgccc	cececectó	cag	

### Figure 12 B

1 agtgaatctg ggcgagtaac acaaaacttg agtgtcctta cctgaaaaat agaggttaga 61 gggatgetat gtgccattgt gtgtgtgtgt tgggggtggg gattgggggt gatttgtgag 121 caattggagg tgagggtgga gcccagtgcc cagcacctat gcactgggga cccaaaaagg 181 agcatettet catgatttta tgtateagaa attgggatgg catgteattg ggacagegte 241 ttttttcttg tatggtggca cataaataca tgtgtcttat aattaatggt attttagatt 301 tgacgaaata tggaatatta cctgttgtgc tgatcttggg caaactataa tatctctggg 361 caaaaatgtc cccatctgaa aaacagggac aacgttcctc cctcagccag ccactatggg 421 getaaaatga gaccacatet gteaagggtt ttgeceteae eteceteeet getggatgge 481 atcettggta ggcagaggtg ggcttcgggc agaacaagcc gtgctgagct aggaccagga 541 gtgctagtgc cactgtttgt ctatggagag ggaggcctca gtgctgaggg ccaagcaaat 601 atttgtggtt atggattaac tcgaactcca ggctgtcatg gcggcaggac ggcgaacttg 661 cagtatctcc acgacccgcc cctgtgagtc cccctccagg caggtctatg aggggtgtgg 721 agggaggget geeceeggga gaagagaget aggtggtgat gagggetgaa teeteeagee 781 agggtgetea acaageetga gettggggta aaaggacaca aggeeeteea caggeeagge 841 ctggcagcca cagtctcagg tccctttgcc atgcgcctcc ctctttccag gccaagggtc 901 cccaggccca gggccattcc aacagacagt ttggagccca ggaccctcca ttctccccac 961 cccacttcca cctttggggg tgtcggattt gaacaaatct cagaagcggc ctcagaggga 1021 gtcggcaaga atggagagca gggtccggta gggtgtgcag aggccacgtg gcctatccac 1081 tggggagggt teettgatet etggeeacca gggetatete tgtggeettt tggageaacc 1141 tggtggtttg gggcaggggt tgaatttcca ggcctaaaac cacacaggcc tggccttgag 1201 tcctggctct gcgagtaatg catggatgta aacatggaga cccaggacct tgcctcagtc 1261 ttccgagtct ggtgcctgca gtgtactgat ggtgtgagac cctactcctg gaggatgggg 1321 gacagaatet gategateee etgggttggt gaetteeetg tgeaateaae ggagaceage 1381 aagggttgga tttttaataa accaettaae teeteegagt eteagtttee eeetetatga 1441 aatggggttg acagcattaa taactacctc ttgggtggtt gtgagcctta actgaagtca 1501 taatatetea tgittaetga geatgageta tgigeaaage etgittigag agettiaigt 1561 ggactaactc ctttaattct cacaacaccc tttaaggcac agatacacca cgttattcca 1621 tccattttac aaatgaggaa actgaggcat ggagcagtta agcatcttgc ccaacattgc 1681 cctccagtaa gtgctggagc tggaatttgc accgtgcagt ctggcttcat ggcctgccct 1741 gtgaatcctg taaaaattgt ttgaaagaca ccatgagtgt ccaatcaacg ttagctaata 1801 ttctcagccc agtcatcaga ccggcagagg cagccacccc actgtcccca gggaggacac 1861 aaacatcctg gcaccctctc cactgcattc tggagctgct ttctaggcag gcagtgtgag 1921 ctcagecca egtagagegg geageegagg cettetgagg etatgtetet agegaacaag 1981 gaceeteaat tecagettee geetgaegge eageacacag ggacageeet tteatteege 2041 ttecaeetgg gggtgeagge agageageag egggggtage aetgeeegga geteagaagt 2101 cctcctcaga caggtgccag tgcctccaga atgtggcagc tcacaagcct cctgctgttc 2161 gtggccacct ggggaatttc cggcacacca gctcctcttg gtaaggccac cccaccccta 2221 ccccgggacc cttgtggcct ctacaaggcc ctggtggcat ctgcccaggc cttcacagct 2281 tccaccatct ctctgagccc tgggtgaggt gaggggcaga tgggaatggc aggaatcaac 2341 tgacaagtee caggtaggee agetgeeaga gtgeeacaea ggggetgeea gggeaggeat 2401 gcgtgatggc agggagcccc gcgatgacct cctaaagctc cctcctccac acggggatgg 2461 teacagagte ceetgggeet teeeteteea eccaeteaet eceteaaetg tgaagaeeee 2521 aggeeeagge tacegteeae actateeage acageeteee etaeteaaat geacaetgge 2581 ctcatggctg ccctgcccca accepttice tggtctccae agccaacggg aggaggccat 2641 gattettggg gaggteegea ggeacatggg eccetaaage cacaccagge tgttggttte 2701 atttgtgeet ttatagaget gtttatetge ttgggacetg cacetecace ettteecaag 2761 gtgccctcag ctcaggcata ccctcctcta ggatgccttt tcccccatcc cttcttgctc 2821 acaccccaa cttgatctct ccctcctaac tgtgccctgc accaagacag acacttcaca 2881 gagcccagga cacacctggg gacccttcct gggtgatagg tctgtctatc ctccaggtgt 2941 ccctgcccaa ggggagaagc atggggaata cttggttggg ggaggaaagg aagactgggg 3001 ggatgtgtca agatggggct gcatgtggtg tactggcaga agagtgagag gatttaactt 3061 ggcagcettt acagcagcag ccagggettg agtacttate tetgggecag getgtattgg 3121 atgttttaca tgacggtctc atccccatgt ttttggatga gtaaattgaa ccttagaaag 3181 gtaaagacac tggctcaagg tcacacagag atcggggtgg ggttcacagg gaggcetgtc 3241 catctcagag caaggettcg tcctccaact gccatctgct tcctggggag gaaaagagca 3301 gaggaccct gcgccaagcc atgacctaga attagaatga gtcttgaggg ggcggagaca 3361 agaccttccc aggctctccc agctctgctt cctcagaccc cctcatggcc ccagcccctc 3421 ttaggeceet caccaaggtg ageteceete cetecaaaac cagacteagt gttetecage 3481 agcgagcgtg cccaccaggt gctgcggatc cgcaaacgtg ccaactcctt cctggaggag 3541 ctccgtcaca gcagcctgga gcgggagtgc atagaggaga tctgtgactt cgaggaggcc

# Figure 12 B (continued)

3601	aaggaaattt	tccaaaatgt	ggatgacaca	gtaaggccac	catgggtcca	gaggatgagg
3661	ctcaggggcg	agctggtaac	cagcaggggc	ctcgaggagc	aggtggggac	tcaatgctga
3721	ggccctctta	ggagttgtgg	gggtggctga	gtggagcgat	taggatgctg	gccctatgat
3781	gtcggccagg	cacatgtgac	tgcaagaaac	agaattcagg	aagaagctcc	aggaaagagt
3841	gtggggtgac	cctaggtggg	gactcccaca	gccacagtgt	aggtggttca	gtccaccctc
3901	cagccactgc	tgagcaccac	tgcctccccq	tcccacctca	caaagagggg	acctaaagac
3961	caccctgctt	ccacccatgc	ctctqctqat	cagggtgtgt	gtgtgaccga	aactcacttc
4021	tgtccacata	aaatcgctca	ctctqtqcct	cacatcaaaq	qqaqaaaatc	tgattgttca
4081	gggggtcgga	agacagggtc	tgtgtcctat	ttatctaaga	gtcagagtcc	tttggagccc
4141	ccagagtcct	gtggacgtgg	ccctaggtag	tagggtgagc	ttggtaacgg	gactagette
4201	ctgagacaag	qctcaqaccc	actctatccc	tagagatcac	ttcagccacc	aggacctgaa
4261	aattgtgcac	acctagaccc	ccttccaagg	catccaggga	toctttccao	tagaggettt
4321	cagggcagga	gaccctctgg	cctgcaccct	ctcttgccct	cagcetecae	ctccttgact
	ggacccccat					
4441	accacagtga	ctttctgcag	gcacatatct	gatcacatca	agteceace	gtgctcccac
4501	ctcacccatg	gtctctcagc	cccagcagcc	ttaactaacc	tetetgatgg	agcaggcatc
4561	aggcacaggc	catagatete	aacgtgggct	aggtagtect	adaccadcad	cadcadacac
4621	agcagcaacc	ctootaccto	attaggaacg	cagaccctct	gccccatcc	teceaactet
4681	gaaaaacact	ggcttaggga	aaggcgcgat	actcaggggt	ccccaaacc	ccccaacccc
4741	agggagtgat	gggactggaa	adadaccasa	taacttaata	agggattcgg	atcattace
4801	tgcagaggct	actataggaa	caaceatca	caacccaaca	agggattegg	categggagaga
4861	gggtgttgct	ccagggacgt	aggacagecg	chagageage	accorateses	ctccagggaga
4921	ggggaggggc	adddadcacc	agetectage	ccgggcgcgg	categggeg	ccatacatat
4981	ttgtctggaa	acceteceet	ccctacca	ctcaccgac	accetaces	accecege
5041	gcccctccgc	acaccaacta	cadageeta	acactacca	ctctctcccc	accegggege
5101	ctggtccaag	cacatcaata	agtgcgttct	acguigueceg	ctagactacc	agecegeeee
5161	cccctcggga	tetetageea	ctgacccct	agacccccgg	atataacaa	coctcaccac
5221	tgcttggtct	taccettaa	gcaccatac	accedent	gegeegeaga	cggcgaccag
5281	atcgacggca	traccaactt	cacteege	taccaccacca	getgegggea	cogcacotto
5341	cagcgcggtg	accadacacac	atacatacta	cecegeageg	gccgggaggg	gggggggggt
5401	tagagagaga	agggggagag	cacetecee	gegggeggeg	tacagagger	ggggccgggc
5461	tgggggcgcg	ctctccacac	cagcigcicg	aggettteet	rgecegeaga	ggtgggcttgg
5521	caacage	actatacaca	teggeggetge	atgeattact	geetagagga	ggcgggccgg
5581	aceastaea	geegegegee	agatagaaa	craggggacg	taataaaa	gtgtcacccc
5641	gcaggcgaga	cacacacac	ggaggtgag	ggaaccacge	cgggcgcggg	gtgggcaggc
5701	attazacett	cacacacaca	ggggcccagg	agggttttta	gggagggagc	gaggaacaga
5761	cccaaaact	ggggcagcgg	teresttte	caacaccggg	tattaataa	gcgcaatcag gtccccgctt
5821	cctccaaaca	gggcgcgccc	ctcgccccc	stastasasa	cocceegge	tggtggctcc
5021	actacagagag	ctcaccatat	ctggggccac	cccccggage	geaageeeag	tgtagcctgg
5001	ctacatttt	ctgagegeat	ctggggcgag	gegegeageg	cocceccea	tgtageetgg
6001	ccttgagga	. ccccgacgcc	cccattata	cattgeatte	tattttaat	ccccttgctt tttatgcatt
6061	ttaatcaaat	ttatatatat	atgaaagtt	occidence	attttaaaa	tcttacgcatt
6121	tcacceacc	ctacacacge	atgaaactt	. dadaallaya	ttttattat	aggtggaccc
6181	ttttaatgt	. geeceeegge	tetteteet	. ctaccaccca	tataaattat	ttcttctaca
6241	atctcccctt	tacttcctct	attttctct	tetegageate	coattatto	gacctctttc
6301	ctctagtttt	attatatat	ctatttccc	tetettese	tttattattta	ctttcaggga
6361	actttcttt	ttttcttt	ttttgagate	coccitigat	stattataa	caggetggag
6421	tacaataaca	tastatasa	torgagace	gageteace	traction	cgattctcct
6481	. cycaacyacy	cccaactac	toggattage	ceegeetee	rggattcaag	gctaattttg
6541	tattttaat	. eccgagrage	tyggactaca tttataaa	t ggcatgegee	teetetee	geraarring
6601	. caddtaata	acctacctta	r acctactact	g ctygtdaagd	tagaccccgaa	ctcctgacct agccaccgcg
6661	caggegated	tteaggean	tttataaa	gigeigggai	cacaggegeg	agecacegeg
6721	ttttagccc	. cccayyydad	, ctcctacaac	. tilataatto	aductititg	cagaaaaaaa
6721	. ceceeggee	trarretter	, cccayacca	a taatteeage	accicgagag	gctgaggtgg
6841	ctatttta	- cyayurugg	ageregagae	. cayectgggc	adcacagtga	gaccetgtet
6001	. coactices	aaaaaytaa	aaaagateta	a dadatttaac		gaaataatta
6961	l aaggeegte	a ggaaguugu	a dayadatgc	gg.gggcct	gergeege	ggtttcctgc
7021	l actacttte	gaaggeeet	s ctactggcag	aduccuagat	. cgrgaggget	ttccttttag
702	l tteggeces	t cagaggacco	toggana	. cryyaggatg	gaagacgctc	acccatggtg
714	l tocatoact	c cayaycayy	a caaaaaaaa	g gagetggtgc	. ctgtgcaggc	tgtggacatt ggcctgaagtc
	Lycatyacti	c cougingues	- yeraayage	a ccacteette	cugaageggg	g geergaagte

### Figure 12 B (continued)

7201 cctagtcaga gcctctggtt caccttctgc aggcagggag aggggagtca agtcagtgag 7261 gagggettte geagtttete ttacaaacte teaacatgee etcecacetg cactgeette 7321 ctggaageee cacageetee tatggtteeg tggteeagte etteagette tgggegeeee 7381 catcacgggc tgagattttt gctttccagt ctgccaagtc agttactgtg tccatccatc 7441 tgctgtcage ttctggaatt gttgctgttg tgccetttee attetttgt tatgatgcag 7501 etcecetget gaegaegtee cattgetett ttaagtetag atatetggae tgggeattea 7561 aggeceattt tgageagagt egggetgaee ttteagecet eagtteteea tggagtatge 7621 getetetet tggcagggag geeteacaaa catgccatge etattgtage agetetecaa 7681 gaatgeteae eteettetee etgtaattee ttteetetgt gaggagetea geageateee 7741 attatgagac cttactaatc ccagggatca cccccaacag ccctggggta caatgagctt 7861 actcttgcca ttgggtggta ctgtttgttg actgactgac tgactgactg gagggggttt 7921 gtaattigta teicagggat tacceccaac agecetgggg tacaatgage etteaagaag 7981 tttaacaacc tatgtaagga cacacagcca gtgggtgatg ctgcctggtc tgactcttgc 8041 cattcagtgg cactgtttgt tgactgactg actgactgac tggctgactg gagggggttc 8101 atagetaata ttaatggagt ggtetaagta teattggtte ettgaaceet geaetgtgge 8161 aaagtggccc acaggctgga ggaggaccaa gacaggaggg cagtctcggg aggagtgcct 8221 ggcaggcccc tcaccacctc tgcctacctc agtgaagttc ccttgtggga ggccctggaa 8281 gcggatggag aagaagcgca gtcacctgaa acgagacaca gaagaccaag aagaccaagt 8341 agateegegg eteattgatg ggaagatgae caggegggga gacageeeet ggeaggtggg 8401 aggegaggea geaceggete gteaegtget gggteeggga teaetgagte cateetggea 8461 gctatgctca gggtgcagaa accgagaggg aagcgctgcc attgcgtttg ggggatgatg 8521 aaggtggggg atgetteagg gaaagatgga egcaacetga ggggagagga geageeaggg 8581 tgggtgaggg gaggggcatg ggggcatgga ggggtctgca ggagggaggg ttacagtttc 8641 taaaaagage tggaaagaca etgetetget ggegggattt taggeagaag ceetgetgat 8701 gggagagggc taggagggag ggccgggcct gagtacccct ccagcctcca catgggaact 8761 gacacttact gggttcccct ctctgccagg catgggggag ataggaacca acaagtggga 8821 gtatttgccc tggggactca gactctgcaa gggtcaggac cccaaagacc cggcagccca 8881 gtgggaccac agccaggacg gcccttcaag ataggggctg agggaggcca aggggaacat 8941 ccaggcagcc tgggggccac aaagtcttcc tggaagacac aaggcctgcc aagcctctaa 9001 ggatgagagg agctcgctgg gcgatgttgg tgtggctgag ggtgactgaa acagtatgaa 9061 cagtgcagga acagcatggg caaaggcagg aagacaccct gggacaggct gacactgtaa 9121 aatgggcaaa aatagaaaac gccagaaagg cctaagccta tgcccatatg accagggaac 9241 gtgatgtcat catcccaccc cattccaggt ggtcctgctg gactcaaaga agaagctggc 9301 ctgcggggca gtgctcatcc acccetcctg ggtgctgaca gcggcccact gcatggatga 9361 gtccaagaag ctccttgtca ggcttggtat gggctggagc caggcagaag ggggctgcca 9421 gaggcctggg tagggggacc aggcaggctg ttcaggtttg ggggaccccg ctccccaggt 9481 gcttaagcaa gaggcttctt gagctccaca gaaggtgttt ggggggaaga ggcctatgtg 9541 ccccacct gcccaccat gtacacccag tattttgcag tagggggttc tctggtgcc 9601 tcttcgaatc tgggcacagg tacctgcaca cacatgtttg tgaggggcta cacagacctt 9661 cacctetcca ctcccactca tgaggagcag gctgtgtggg cctcagcacc cttgggtgca 9721 gagaccagca aggcctggcc tcagggctgt gcctcccaca gactgacagg gatggagctg 9781 tacagaggga gecetageat etgecaaage cacaagetge tteeetagea ggetgggge 9841 tectatgeat tggeceegat etatggeaat ttetggaggg ggggtetgge teaactettt 9901 atgecaaaaa gaaggeaaag catattgaga aaggeeaaat teacatttee tacageataa 9961 tetatgecag tggccccgtg gggcttggct tagaattece aggtgctett cccagggaac 10021 catcagtetg gactgagagg acettetete teaggtggga eceggeeetg teeteeetgg 10081 cagtgccgtg ttctgggggt cctcctctct gggtctcact gcccctgggg tctctccagc 10141 tacctttgct ccatgttcct ttgtggctct ggtctgtgtc tggggtttcc aggggtctcg 10201 ggcttccctg ctgcccattc cttctctggt ctcacggctc cgtgactcct gaaaaccaac 10261 cagcatecta eccetttgga ttgacacetg ttggecacte ettetggeag gaaaagteae 10321 cgttgatagg gttccacggc atagacaggt ggctccgcgc cagtgcctgg gacgtgtggg 10381 tgcacagtct ccgggtgaac cttcttcagg ccctctccca ggcctgcagg ggcacagcag 10441 tgggtgggcc tcaggaaagt gccactgggg agaggctccc cgcagcccac tctgactgtg 10501 ccctctgccc tgcaggagag tatgacctgc ggcgctggga gaagtgggag ctggacctgg 10561 acatcaagga ggtcttcgtc caccccaact acagcaagag caccaccgac aatgacatcg 10621 cactgctgca cctggcccag cccgccaccc tctcgcagac catagtgccc atctgcctcc 10681 cggacagcgg ccttgcagag cgcgagctca atcaggccgg ccaggagacc ctcgtgacgg 10741 getggggeta ceacageage egagagaagg aggeeaagag aaacegeace ttegteetea 10801 acttcatcaa gattcccgtg gtcccgcaca atgagtgcag cgaggtcatg agcaacatgg

## Figure 12 B (continued)

10861	tgtctgagaa	catgctgtgt	gcgggcatcc	tcggggaccg	gcaggatgcc	tgcgagggcg
		gcccatggtc				
10981		gggctgtggg				
11041		gatccatggg				
11101	cttagcgacc	ctccctgcag	ggctgggctt	ttgcatggca	atggatggga	cattaaaqqq
11161	acatgtaaca	agcacaccgg	cctgctgttc	tgtccttcca	tccctcttt	gggctcttct
11221	ggagggaagt	aacatttact	gagcacctgt	tgtatgtcac	atgccttatg	aatagaatct
11281	taactcctag	agcaactctg	tggggtgggg	aggagcagat	ccaagttttg	cggggtctaa
11341		gttgaggggg				
11401		gccttttcca				
11461	gtgaggcttg	accagettte	cagctagccc	agctatgagg	tagacatgtt	tagctcatat
11521	cacagaggag	gaaactgagg	ggtctgaaag	gtttacatgg	tggagccagg	attcaaatct
11581	aggtctgact	ccaaaaccca	ggtgcttttt	tctqttctcc	actotcctoo	aggacagetg
		gctcagtgtg				
		ggttcagccc				5 5

## Figure 13 (A)

#### SEQ ID NO:3

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagccctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tetecactat ggeaactaae teacteaatt tteeeteett ageageatte catetteeeg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet aetetateae aaggeeagta eeacaeteat gaagaaagaa eacaggagta getgagagge taaaacteat caaaaacact acteettte etetaceeta tteetcaate ttttacettt tecaaateec aateeccaaa teagtttte tetteetae teeetetee ccttttaccc tccatggtcg ttaaaggaga gatggggagc atcattctgt tatacttctg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccatte taagagettg tatggttatg gaggtetgae taggeatgat tteaegaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagtigte etittetggt tiegigtiea ceatggaaca tittgattat agttaateet tctatcttga atctt

#### SEQ ID NO:76

ggcctctg actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg cttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagccctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tetecaetat ggcaactaae teaeteaatt tteeeteett agcageatte catetteeeg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet actetateae aaggeeagta ceacacteat gaagaaagaa cacaggagta getgagagge taaaacteat caaaaacact acteettte etetaceeta teetcaate ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctctc cettttaccc tecatggtcg ttaaaggaga gatggggage atcattetgt tataettetg tacacagtta tacatgteta teaaaceeag aettgettee atagtggaga ettgetttte agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

## Figure 13 (B)

#### SEQ ID NO:77

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cocagcoctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tctccactat ggcaactaac tcactcaaat ttccctcctt agcagcattc catcttcccg atettetttg etteteeaac caaaacatea atgtttatta gttetgtata cagtacagga tctttggtct actctatcac aaggccagta ccacactcat gaagaaagaa cacaggagta getgagagge taaaacteat caaaaacact acteettte etetaceeta tteeteaate ttttacettt tecaaateec aateeccaaa teagtttte tetteettac teeetete ccttttaccc tccatggtcg ttaaaggaga gatggggagc atcattctgt tatacttctg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccatte taagagettg tatggttatg gaggtetgae taggeatgat tteacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

#### SEQ ID NO:78

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cecagecett gacaaaattg tgaagttaaa ttetecaete tgtecateag atactatggt tetecaetat ggcaactaae teaeteaatt tteeeteett ageageatte catetteeeg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet aetetateae aaggeeagta eeacaeteat gaagaaagaa cacaggagta gctgagaggc taaaactcat caaaaacact actccttttc ctctacccta ttcctcaatc ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctctc cettttacce tecatggteg ttaaaggaga gatggggaeg atcattetgt tataettetg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgctttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatata agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccatte taagagettg tatggttatg gaggtetgae taggeatgat tteacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

## Figure 13 (C)

#### SEQ ID NO:79

### SEQ ID NO:80

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg cttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagccctt gacaaaattg tgaagttaaa ttetecaete tgtecatcag atactatggt tetecaetat ggeaactaac teaeteaatt tteeeteett ageageatte catetteeeg atttettettig etteteeaac caaaacatea atgittatta gitetgiata cagiacagga tetttggtet aetetateae aaggeeagta eeacaeteat gaagaaagaa eacaggagta gctgagaggc taaaactcat caaaaacact actcetttte etetacecta tteetcaate ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctctc cettttacce tecatggteg ttaaaggaga gatggggage atcattetgt tataettetg tacacagtta tacatgteta teaaacceag acttgettee atagtggaga ettgettte agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg caagattggc atatcattgt aactaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

## Figure 13 (D)

#### SEQ ID NO:81

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagcctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tctccactat ggcaactaac tcactcaatt ttccctcctt agcagcattc catcttcccg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet aetetateae aaggeeagta eeacaeteat gaagaaagaa cacaggagta getgagagge taaaacteat caaaaacact acteettte etetaceeta tteeteaate ttttaccttt tecaaateee aateeceaaa teagttttte tetttettae teeetetete ccttttaccc tccatggtcg ttaaaggaga gatggggagc atcattctgt tatacttctg tacacagtta tacatgicta tcaaacccag actigcitcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa etgetaagttgte ettttetggt ttegtgttea ceatggaaca ttttgattat agttaateet tctatcttga atctt

#### SEQ ID NO:82

ggcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagccett gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tetecaetat ggeaactaae teaeteaatt tteeeteett ageageatte eatetteeeg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet actetateae aaggeeagta ceacacteat gaagaaagaa cacaggagta getgagagge taaaacteat caaaaacact acteettte etetaceeta tteeteaate ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctctc cettttacce tecatggteg ttaaaggaga gatggggage atcattetgt tatacttetg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatata aatatatat taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttatcct tctatcttga atctt

### Figure 13 (E)

#### SEQ ID NO:83

gcctctc actaactaat cactttccca tcttttgtta gatttgaata tatacattct atgatcattg ctttttctct ttacagggga gaatttcata ttttacctga gcaaattgat tagaaaatgg aaccactaga ggaatataat gtgttaggaa attacagtca tttctaaggg cccagcctt gacaaaattg tgaagttaaa ttctccactc tgtccatcag atactatggt tctccactat ggcaactaac tcactcaatt ttccctcctt agcagcattc catcttccg atcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga tetttggtet actetateae aaggeeagta ceacacteat gaagaaagaa cacaggagta gctgagaggc taaaactcat caaaaacact actccttttc ctctacccta ttcctcaatc ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctctc ccttttaccc tccatggtcg ttaaaggaga gatggggagc atcattctgt tatacttctg tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgcttttc agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag ttaagttatt ttatatatat aatatatata taaaatatat aatatacaat ataaatatat agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aagtcattcc aatcagccaa ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct tctatcttga atctt

# Figure 14

AATTCTGTA AGCATTTCCT ATGTGTACCT GCCCCTGGGC AAGGTGGGCC TGACTTGTTA	-1403
AGTGTTAGA GTTTTACCCT GTTCCTCTAG GAGGGCCTGG TACCACCACA GCCCAGCATG	-1343
TGTGGTGCC TCAGCAGGAG GCATCTGGTT ACAATCAACA CAAGCTGTTC CAGCCAATTT	-1283
NAGARACTT CASGAGGART AGGGTTTTAG GRGGGCATGG GGACCCTCCT GCACCCGAAG	-1223
CAGGATGTG CCACCAATCA TAAGGAGGCA GGGGCCTCCT TCCGCTGCTC CCTGGGACTC	-1163
CTAGGTGTC CGTGGCCTCA GCCCCCCTCT GCACACCTGC ATCTTCCTTC TCATCAGCTT	-1103
CTCTGCTTT AAGCGTAAAC ATGGATGCCC AGGACCTGGC CTCAATCTTC CGAGTCTGGT	-1043
acttanggig tactgacagt gigagaccct actcctctga tcaatcccct gggttggiga	-983
TTTCCTTGTG CAATCAATGG AAGCCAGCGA GGCAGGGTCA CATGCCCCGT TTAGAGGTGC	-923
rgactiggag arggarcetg ggcaretcit cccaggarca ggtagggcag ggaggrargg	-863
GGGCCATCTC TGGTGCAGCC CGGTTCGGAG CAGGAAGACG CTTAATAAAT GCTGATAGAC	-803
TGCAGGACAC AGGCAAAGGT GCTGAGCTGG ACCCTTTATT TCTGCCCTTC TCCCTTCTGG	-743
CACCCCGGCC AGGAAATTGC TGCAGCCTTT CTGGAATCCC GTTCATTTTT CTTACTGGTC	-683
CACAAAAGGG GCCAAATGGA AGCAGCAAGA CCTGAGTTCA AATTAAATCT GCCAACTACC	-623
AGCTCAGTGA ATCTGGGCGA GTAACACAAA ACTTGAGTGT CCTTACCTGA AAAATAGAGG	-563
TTAGAGGGAT GCTATGTGCC ATTGTGTGTG TGTGTTGGGG GTGGGGATTG GGGGTGATTT	-503
GTGAGCAATT GGAGGTGAGG GTGGAGCCCA GTGCCCAGCA CCTATGCACT GGGGACCCAA	-443
ARAGGASCAT CTTCTCATGA TTTTATGTAT CAGARATTGG GATGGCATGT CATTGGGACA	-383
GCGTCTTTTT TCTTGTATGG TGGCACATAA ATACATGTGT CTTATAATTA ATGGTATTTT	-323
AGATTIGACG AANTAIGGAA TATTACCIGT IGTGCIGAIC ITGGGCAAAC TATAATAICI	-263
CTGGGCAAAA ATGTCCCCAT CTGAAAAACA GGGACAACGT TCCTCCCTCA GCCAGCCACT	-203
ATGGGGCTAA AATGAGACCA CATCTGTCAA GGGTTTTGCC CTCACCTCCC TCCCTGCTGG	-143
ATGGCATECT TGGTAGGCAG AGGTGGGCTT CGGGCAGAAC AAGCCGTGCT GAGCTAGGAC	-83
CAGGAGTGCT AGTGCCACTG TTTGTCTATG GAGAGGGAGG CCTCAGTGCT CAGGGCCAAG	-23
CANATATITE TESTTATEGA TIAACTCGAN CTCCNEGCTE TCATEGCEGC NEGACGECGN	+38
ACTTGCAGTC TCTCCACGAC CCGCCCCTGT GAGTCCCCCT CCAGGCAGGT CTATGAGGGG	+98
TETEGAGGA GGGCTGCCCC CGGGAGAAGA	
1350 bp	
. Wom man civi if	17 -39

MET TRP GLN LEU -39
----- AAGAAGTOOT COTCAGACAG GTGCCAGTGC CTCCAGAATG TGG CAG CTC +1527

Figure 15

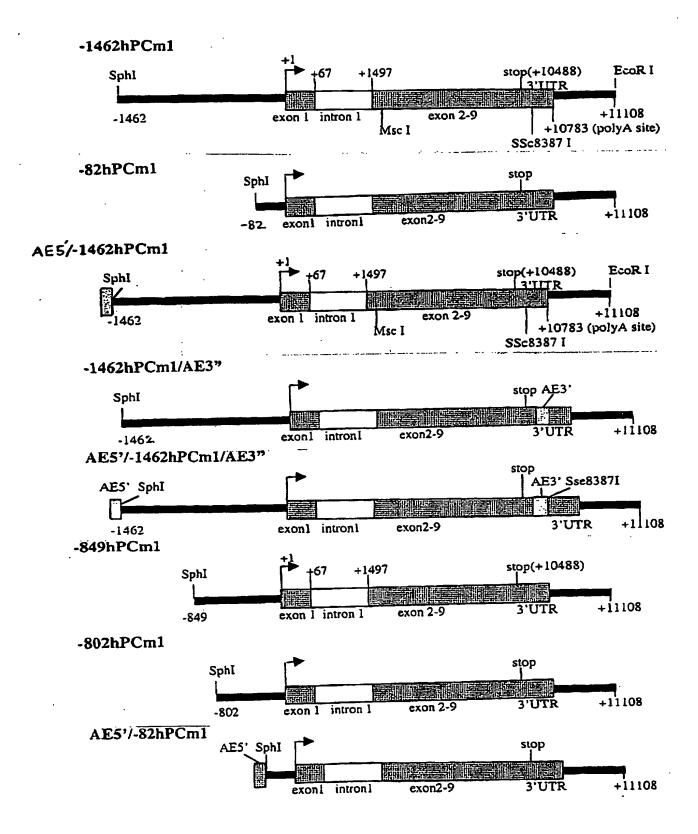


Figure 16

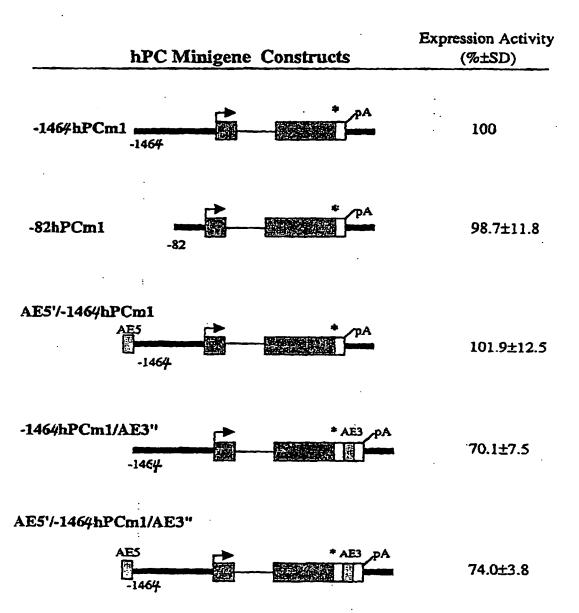


Figure 17A

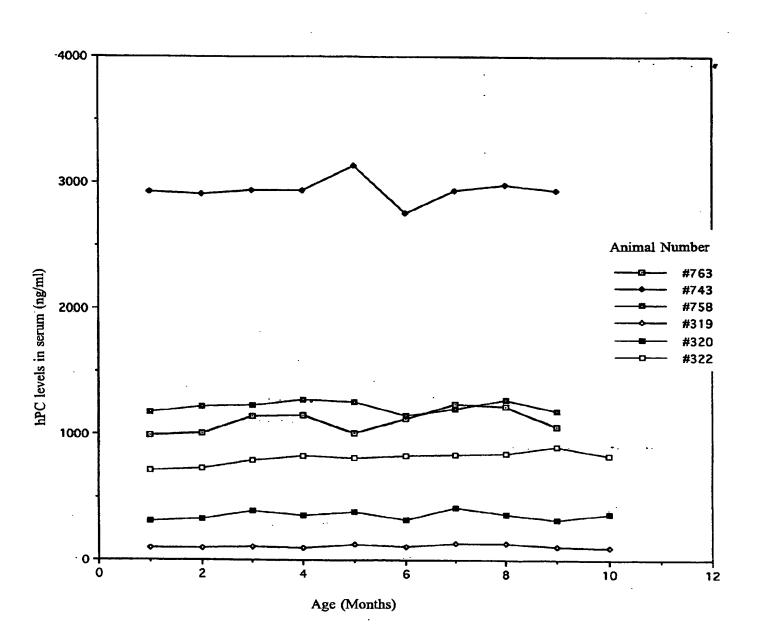


Figure 17B

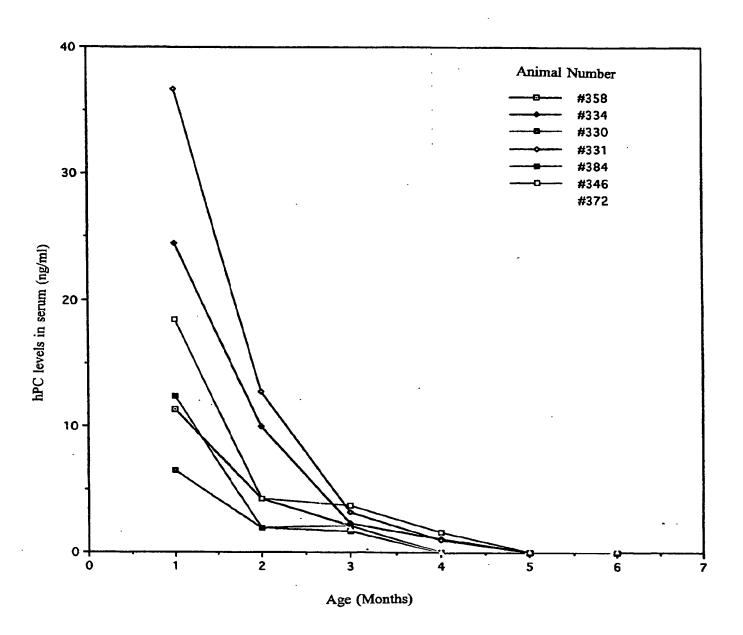
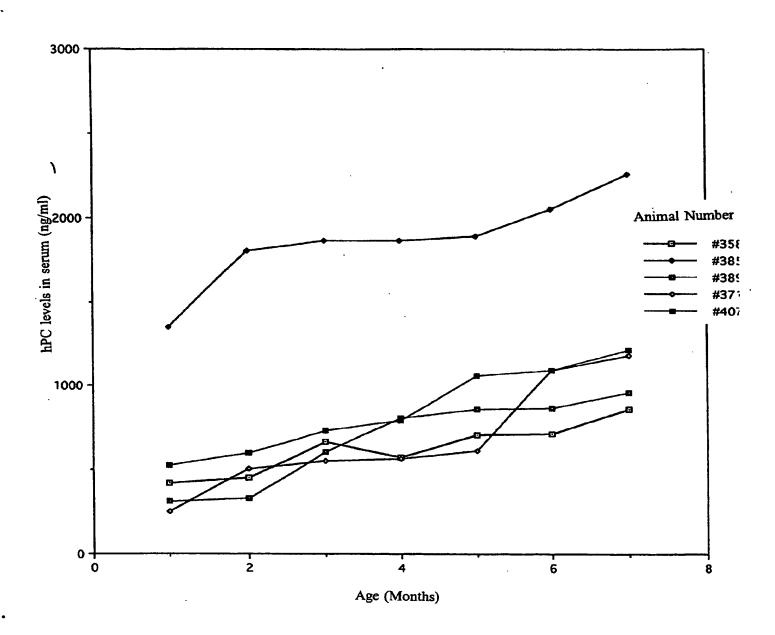


Figure 17C



### SEQUENCE LISTING

5	Kurachi, Kotoku  Kurachi, Sumiko	
	<120> Nucleotide Sequences for Gene Regulation and Methods of Use Thereof	
10	<130> UM-04305	
10		
	<140> xx	
	<141> 2000-06-06	
	<150> 09/328,925	
15	<151> 1999-06-09	
	<160> 208	
20	<170> PatentIn Ver. 2.0	
	<210> 1	
	<211> 32	
	<212> DNA	
	<213> Homo sapiens	
25		
	<400> 1	
	agccattcag tcgaggaagg atagggtggt at	32
	<210> 2	
30	<211> 7	
	<212> DNA	
	<213> Homo sapiens	-
	<400> 2	
35	gaggaag	7
	<210> 3	
	<211> 1273	
	<212> DNA	
40	<213> Homo sapiens	



<400> 3 ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60 gatcattqct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120 gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180 5 cagecettga caaaattgtg aagttaaatt etceactetg tecateagat actatggtte 240 tccactatgg caactaactc actcaatttt ccctccttag cagcattcca tcttcccgat 300 cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360 tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtaqc 420 tgagaggcta aaactcatca aaaacactac tccttttcct ctaccctatt cctcaatctt 480 10 ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540 ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600 cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660 aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720 aagttatttt atatatata tatatata aaatatataa tatacaatat aaatatatag 780 15 tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840 agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900 agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960 ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020 ctagtagaga ctttgaggaa gaattcaaca gtqtgtcttc agcagtgttc agagccaagc 1080 20 aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140 ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200 aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttaatccttc 1260 tatcttgaat ctt 1273 25 <210> 4 <211> 38059 <212> DNA <213> Homo sapiens 30 <400> 4 gtatatctag aaaaccccat tgtctcattc caaaatcacc ttaagatgga taggcaactt 60 cagcaaagtc tcaggataac aaaatcaatg tgcaaaaatc acaggcattc ttatacacca 120 atagcagaca aacagacagc caaatcatga gtgaactccc attcacaatt gcttcaaaga 180 gaataaaata cctaggaatc ctacttacaa gggatgtgaa ggacctcttc aaggagaact 240 35 acaaaccact gctcaatgaa ataaaagagg atacaaacaa atggaagaac attacatgct 300 catgggtagg aagaatcaat atcatgaaaa tggccataat gcccaaggta atttatagat 360 tcaatgccat ccccatcaag ctaccaatga ctttcttcac agaattggaa aaaactactt 420 taaagttcat atggaaccaa aaaagagccc gcatcgccaa gtcaatccta agccaaaaga 480 acaaagctgg aggcatcatg ctacctgact tcaaactata ctacaaggct acagtaacca 540 40 aaacagcatg gtactggtac caaaacagag atacagacca atggaacaga acagagccct 600 cagaaataat gccacatatc tacaactatc tgatctttga aaaacctgac aaaaacaaga 660 aatggggaaa ggaatcccta attaataaat ggtgctggga aaactggcta gccatatgta 720 gaaagctgaa actggatccc ttccttatac cttatacaaa aattaattca agatggatta 780 aagacttcat tgttagacct aaaaccataa aaaccctaga agaaaaccta ggcaatacca 840

ttcaggacat aggcatgggc ttggacttca tgtctaaaac accaaaagca atggcaacaa 900

45



	aagccaaaat	tgacaaatgg	gatcaaatga	aactaaagag	cttctgcaca	gcaaaagaaa	960
	ctaccatcag	agtgaacagg	caacctaaag	aatgggagaa	aatttttgca	atctactcat	1020
	ctgtccaagg	gctaatatct	agaatctaaa	atgaactcaa	acaaatttac	agaaaaaac	1080
•	aaacaacccc	atcaacaagt	gggtgaagga	tatgaacaga	cacttctcaa	aagaagacat	1140
5	ttatgcagcc	aacagacaca	tgaaaaaatg	ctcagcatca	ccggccatca	gagaaatgca	1200
	aatcaaaacc	acaatgagat	accatctcac	acaagttaga	atggcgatca	tcaaaaactc	1260
	aggaagcaac	aggtgctgga	gaggatgtgg	agaaatagga	acactttgac	actgttggtg	1320
	ggactgtaaa	ctagttcaac	cattgtggaa	gtcagtgtgg	cgattcctca	gggatctaga	1380
	cctagaaata	ccatctgacc	cagccatccc	attattgggt	atataccaaa	gtattataaa	1440
10	tcatgctgct	ataaagacac	atgcacacgt	atgtttattg	cggcactttt	cacaatagca	1500
	atgacttgga	accaacccaa	atgtccaaca	atgatagact	ggattaagaa	aatgtggcac	1560
	atatacacct	aggaatacta	ggcagccata	aaaagaaaat	gagttcatgt	cctttgtagg	1620
	gcatggatga	agctagaaac	catcattctc	agcaaactat	cgcaaggaca	aaaaaccaaa	1680
	caccgcatgt	tctcactcat	aggtgggaac	tgaacaatga	gaacacttgg	acacaggaag	1740
15	gggaacatca	cacaccgggg	cctgttgtgg	ggtggggggc	gaggggaggg	atagcattag	1800
	gggatatacc	taatgctaaa	tgacgagtta	atgggtacag	cacaccaaca	tggcacatgt	1860
	atacatatgt	aacaaacctg	ctcgttgtgc	acatgtaccc	taaaacttaa	agtataataa	1920
	taaaaaaaag	atcattctaa	aatttataca	agcccttaga	acagttaaaa	atatcttacc	1980
	aaaagaagaa	taaagttgga	ggaatcactc	tacctaatat	aaagtcttac	tacatagcta	2040
20	cagtaattat	gacagtgtta	tattggcaga	gggataaata	catcaatggc	acaaagaata	2100
•	gatagagaaa	ctggaagtag	acccaaaaca	atatggttaa	ctgacttacg	aaaaaatttc	2160
	agaagccatt	cagtcgagga	aggatagggt	ggtattgttg	ttttttgttt	taacaaattg	2220
	tgctggataa	attggacata	cctatggaaa	aaaaaatgaa <sup>.</sup>	gtttgaccta	aacatcatac	2280
	tttacacaaa	tattaactca	aaatggagca	tgggcataaa	tctaaaactt	caaactgtaa	2340
25	aacatttaga	aaaaaatagg	aaaaaacta	tcaggatcta	gtgttagtgg	aagagttcta	2400
						äaaattctac	
						caatatattt	
						aagaatctga	
20	_	_				gaaatgtaaa	
30						gggtcccact	
						tgaagatgtg	
		•		·		atcagaggtg	
						ttggaaatag	
25						gctcagcttg	
<b>35</b> .				•		ggttatgcag	
					•	tttaggatat	
						gtatgcttgc	
			•		_	acatgatttg	
40						gttctttgtt	
+∪						ttatgatgca	
						gttagaaaac	
						cagaatatag	
						aaaatgggaa	
45						tatatgacat	
73	cccaaggcct	cayaaytatg	caacyayytg	Lycocotaat	LLLLLAAACE	atatatcttc	2000



	aatttaaagt	tttagttaaa	acataaagat	taacctttca	ttagcaagct	gttagttatc	3660
	accaaagctt	tţcatggatt	aggaaaaaat	cattttgtct	ctatgtcaaa	catcttggag	3720
	tttgattatt	tggggaaaca	caatactcag	ttgagttccc	taggggagaa	aagcaagctt	3780
	aagaattgac	ataaagagta	ggaagttagc	taatgcaaca	tatatcactt	tgtttttca	3840
5	caactacagt	gactttatgt	atttcccaga	ggaaggcata	cagggaagaa	attatcccat	3900
	ttggacaaac	agcatgttct	cacaggaagc	atttatcaca	cttacttgtc	aactttctag	3960
	aatcaaatct	agtagctgac	agtaccagga	tcaggggtgc	caaccctaag	caccccaga	4020
	aagctgactg	gccctgtggt	tcccactcca	gacatgatgt	cagctgtgaa	atcagactga	4080
	aatgctgaaa	taacgataaa	aaaaaataca	gaggttaaac	tagcaaagtg	agtaaagtca	4140
10	agggataaag	aaaatttgtt	ggaaaactca	caaagcagga	cataaagcaa	ggccattaga	4200
	tatatctcat	tagtgtgaca	tctgggagga	caaagcatcc	aaaccctttc	ttctatataa	4260
	gtggtgagat	gatgaaggtt	gtaagaggct	tetgeceet	tgaagacttc	agatgctggg	4320
	gaaaggatag	ataagaataa	ggatgaacct	ggcttttgga	gcctgggaaa	taatgactag	4380
	cgataaacct	gaagggaagt	taagtatacg	atccccagat	aatactaagg	agaaaggcaa	4440
15	tgtgattctg	cagccattgt	agccagagat	aataagccct	tgaggaaggg	gccaggggaa	4500
	tttttctaag	gatagacagt	attaatgcag	cactctcttc	tgctattaaa	ctctcattgg	4560
	cttctaaaag	gagtttcggt	gagtgatttg	ctgagatgtt	tgcattttca	tgctgctgcc	4620
	tttaggttat	tattgcaaca	gtttggaatt	ttgaaattaa	aacagttctg	taaaaccagt	4680
			catcaaagat				
20	ctacggtgcc	aggtactgtg	tcagggtact	aggggtatgg	ggataaacca	gactccctct	4800
			gccaggtgag				
			agaaaggaac				
			ttaaaagatg				
0.5			agaaggatca				
25			gtatggttcc				
			gataatgaag				
			gtataaacat				
			aatgaaatga				
20			attgcaatta				
30	•		gaggggagtt				
			gaagggaaa				
			taatgtctaa				
			agtatgttac				
35			aagatgacag	_			
33			agttactgag				
			ctacattatg			_	
			tgactatagg				
			gggcatctgg ttgtatgctc				
40			gcctgtaggg				•
			ccctacatgc				
			aaaggaaaaa				
			gaacaggagt			_	
			accagaatag			-	
45			aaattattta				
-	3333~3~6			2-2249	a	uucccaaaat	9500

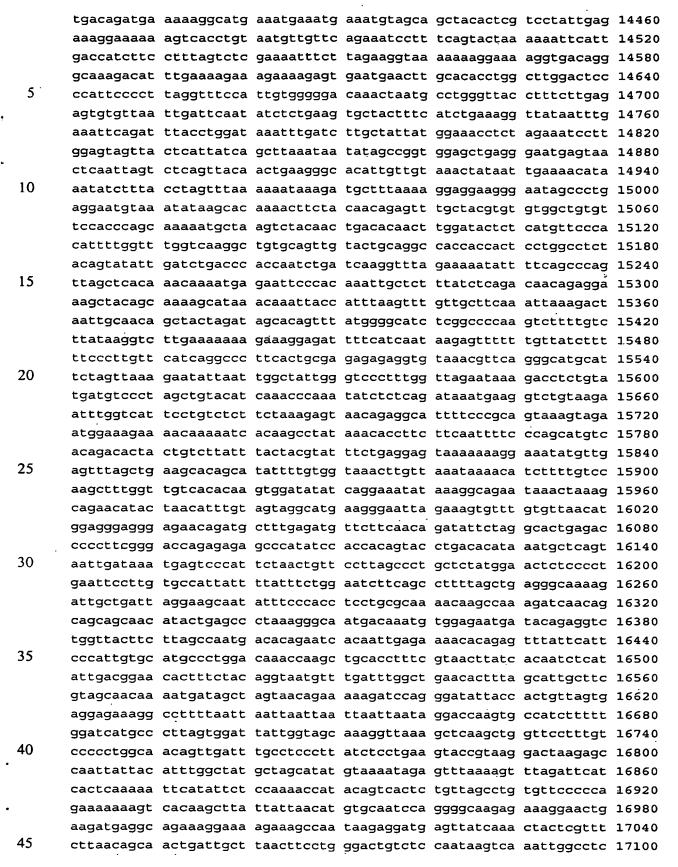


	tattctttag	gtcattcctg	attgtgacaa	tagtcatttc	attatataaa	tgtgattaag	6360
	gaaggaaaag	agctacacag	aagttattaa	agagctaaag	agaattgaga	aatttaaaac	6420
•	agaagaaagt	agggccaaca	tgaaaggagt	agggagaaaa	agagataacc	agcatattgt	6480
	ctcactgatc	ctgccaacac	ctgtgagata	gatattctta	ttatactact	gttaaaccta	6540
5	atttacattt	gacaaagtta	aggttcagag	cttgtgtgac	ttgtccaagg	tcacaggtct	6600
1	agaggaggca	gatacttgat	tcaaacctat	ttctgtctga	tctgattcta	aagtctgttt	6660
	tttcactcaa	ccacactgta	cagtcagctc	tccttgtgag	ttccacagcc	acagattcaa	6720
ı	ttaactgcag	atcaaaaata	ttcaagaaaa	aaatggatgg	ttcgatctct	actgaacatg	6780
	tacagactct	tttatctttc	attattccct	aaacaataca	gcataacaac	tatttacata	6840
10	gcatttacat	tgtattagct	attaagagaa	acctagagat	gatttaaagt	acaaaggagg	6900
	atgtgtttag	gttatatgca	aatagtaagc	catttttata	tcggagactt	gagcatccac	6960
	agatcttgat	atttgcaggg	ggtcttgcca	ccaattttcc	atggatactg	aggaacgact	7020
	gtaaatggat	gcaggcatgg	atgctattta	ggagtgtcca	gggccaagta	aatgagttgc	7080
•	tgagcagaga	ggtgggtgga	ggctgtgagg	catcaatatg	tggtggcatc	catctgcatt	7140
15	ttggtgattt	ttttccttca	cagtcctcgg	ctgtctggga	agagaaggat	gaaggcagat	7200
	ggctgctcca	atttaggggc	taggattgca	gggtgggcac	agcattgcaa	acgagtgaag	7260
	gaaattgaga	aatatggcca	atgaagagtt	gaagagaggc	ctggcatggt	ggctcacacc	7320
	tataatccca	gcactttcag	aggcccaggc	aggcagatca	cttgaggtca	ggagttcgac	7380
	accagcctgg	ccaacaaggt	gaaatggtga	aaccccggct	ttactaaaaa	tacaaaaatt	7440
20	agctgggcat	ggtggcgggt	gcctgtaatc	ccagctactt	gggaggctga	ggcaggagaa	7500
	tagcttgaac	ctgggagatg	gaggttgcag	tgagctgaga	tcgcaccact	gcactccagc	7560
	ctgggcgaca	gagcaagact	ctgtcaaaaa	aaaaagagtt	gaagagaaaa	agtctaggct	7620
						aagcagggat	
0.5						ggaagaaagt	
25						tttttggagt	
						tggtcctctt	
						ggcaacttct	
						tgaaaacaca	
20						attctgagaa	
30	_			_		acctaaacct	
						tcctaggctc	
				_		tggtaacagt	
				- <del>-</del>	-	taaccttatg	
35						agcaaatgat	
33						tgtatatatg	
						tatgtacaca	
	•					tatatgtata	
•						gagagagaga	
40						gagacatcta	
			_	<del></del> -		aaaaacatat	
				•		ccaaccaaag	
•			•			taacatctag tcttcaggag	
						agggtgatgg	
45						tgaagttaac	
•	uccaeccego.	adayattett	aacyayctat	ccccaagcga	ryacaaayrg	LyaayttaaC	9000

	cgctcaattt	gagttctttc	tttttcatcc	aaagtaaatt	caaatatgat	tagaaatctg	9060
	accttttatt	actggaattc	tcttgactaa	aagtaaaatt	gaattttaat	tcctaaatct	9120
	ccatgtgtat	acagtactgt	gggaacatca	cagattttgg	ctccatgccc	taaagagaaa	9180
	ttggctttca	gattatttgg	attaaaaaca	aagactttct	taagagatgt	aaaattttca	9240
5	tgatgttttc	ttttttgcta	aaactaaaga	attattcttt	tacatttcag	tttttcttga	9300
	tcatgaaaac	gccaacaaaa	ttctgaatcg	gccaaagagg	tataattcag	gtaaattgga	9360
	agagtttgtt	caagggaacc	ttgagagaga	atgtatggaa	gaaaagtgta	gttttgaaga	9420
	agcacgagaa	gtttttgaaa	acactgaaag	aacagtgagt	atttccacat	aatacccttc	9480
	agatgcagag	catagaatag	aaaatcttta	aaaagacact	tctctttaaa	attttaaagc	9540
10	atccatatat	atttatgtat	gttaaatgtt	ataaaagata	ggaaatcaat	accaaaacac	9600
	tttagatatt	accgttaatt	tgtcttcttt	tattctttat	agactgaatt	ttggaagcag	9660
	tatgttggta	agcaattcat	tttatcctct	agctaatata	tgaaacatat	gagaattatg	9720
	tgggttttt	ctctgcataa	atagataata	tattaaactt	tgtcaaaagg	actcagaaag	9780
	atcagtccaa	ccctctaacc	catattggat	ggtgatatac	tacagggtta	tgccagtgtg	9840
15	ggaactatcg	ctggtaaata	agtttaatcc	tccctagggc	ttcacaaaga	acattgttcc	9900
	accccaggag	ggtggaagga	agaaactgaa	atgattgtgt	cttagaacct	aatgaaagtt	9960
	tgcattcctc	agtaaaatca	gagactgctg	attgacttaa	atgtttatag	cttcaaagtc	10020
	ctcctcatta	tcatggccca	gaagcccttc	catgattgtc	cttccccacc	ctccccatta	10080
	cccttcttgc	ctcctctgct	acttctctcc	tcgcacactg	ggctccagcc	accctggcct	10140
20	tcctgtcact	tcttgcacac	tctaggaatg	ctcccacttt	ggaggcttta	tctggctgtt	10200
	tctcttattt	ggctgttccc	aacttcctgt	gggctgactc	cctcacctcc	ttcgggtctt	10260
	tgcccaaatg	ttaccatctt	aatgaggcct	accttcacca	tctattaata	cttcaacctg	10320
•	ccccagtagc	cttaccactc	tagacacctg	tacagaactc	cactctactt	tttaacagag	10380
	cttttcacca	tctaatgtat	catataattt	cttactcata	ctatttatca	tttattttct	10440
25	cctactccac	taaaatgcaa	gtttcatgtt	ggcagggata	ttcaattgtt	ttgtttattg	10500
	atatattcct	agcacctaga	acagtatctg	gaaaagaggt	actcagtaaa	tatttatcaa	10560
	atgaattgac	caaaagaagg	aaaactcaaa	actttaatga	caactaactt	taaagctaca	10620
	ataacttaaa	attcagagta	ggattttgag	ggagggtaag	tttcaaagat	tgacttacct	10680
	aagactatct	gcataaataa	aaagaaatta	atccagacaa	caaattcacc	aacttccatc	10740
30	aattggaaat	ccaattctat	tttctacagt	ttatgttctg	gagacactac	tggacactct	10800
						gaaaaagagt	
	tgactctgtt	atattgtttt	atctaccttt	ccttgatctt	agaaacgaat	actaccatac	10920
	cagcttctac	tgaggtgccc	cctaaagtta	gtccaaatag	gtctttgcaa	tctccattcc	10980
			_		-	atgccgattt	
35	tccccaccaa	adaattcctg	actattaaac	tcctacaatc	ccttcattgc	tcactcccca	11100
	•					gtacactgtt	
	tgctatacca	caggtatagc	tatctggaaa	acatggaggg	tattattctg	ttactactgc	11220
			_	_		cctgcctctt	
40						tcttgcacat	
. 40				= .		tcacataaag	
						ctttccagtt	
`						attattaact	
					_	aaaggtctcc	
45						aaccccttgc	
45	tttcactatt	acacctcagg	gccttcaaat	ctaaatgcag	ttattcatta	aacaggaacc	11700

	tggtagtctt	aaacaggaat	ctctcacttg	gtaagatctt	gtctcttgtt	gtatttgacc	11760
	ccaactgtct	atggctttgc	ctgaacccaa	agtacacaca	gcctagaaac	caaaggagaa	11820
	ccaaatgtgg	gataaaatga	cactcatttt	aacgacatgt	ctcagcaaat	gagttcctgt	11880
	gtagctggct	gaaagcccag	accctttcag	taaaacatcc	tgaataattç	acatttgttg	11940
5	gtctataata	taaagggcaa	atgtagctca	tttttagacc	agttctgaac	atcaatagta	12000
	acaaaccaga	gataaccgat	tttgttttca	tagaattgga	acaaattaga	gtatctgtgc	12060
	aaaagcatat	cagatctagg	agcagagggg	acaaggtcta	atttttaaat	aagcaaattt	12120
	tccagggagg	gactacttat	gataaaggga	tattagtctc	ttagtcaacg	gaacctggat	12180
	acacgcttct	gacagagaag	agggagaata	ggcaggaatc	tacacaccag	atgtcaagga	12240
10	gatttgcttt	aaaatacgac	tgataattag	aaatttctca	gtttcccct	tttccctcat	12300
	tctttgattc	ttattgttat	ctttatctct	tactcctttg	tttctcatat	attgagtctt	12360
	acagatcaag	ctcccatttt	tttcttcagg	ggtattttc	tagttcaaag	tgcctaccat	12420
	ctcccttctg	gttctattca	tccttctctc	ccaaagctcc	tttagaagtg	tggattaagg	12480
	cagagcacta	agaaaccaga	cttaaagatt	cccttctcat	tctgactttt	ctcctttcac	12540
15	ctattccttc	ctcctgtttt	cttaccatca	gtgtcttcaa	aggctttcaa	gtacacggta	12600
	aatgcagaaa	cttcaagaaa	ggcagaatgg	aaacataacc	aatgcataca	taaataaagc	12660
	acactgtaga	atctttttaa	attctgtatg	atatatcgaa	tgctgtctct	cacattacct	12720
	agaccatttg	aaaccgaatt	tgtaaaacat	agactatctt	taagtagtaa	cagatgcttc	12780
	tgacatgttt	tċtattgtct	tgaaccatta	ctgcatatga	tacatcaaag	ttaagtgaca	12840
20	atacaagaaa	gcagattcat	ttgctccctg	cctaggccgt	cagttcctaa	agtggaaacg	12900
	ccatatatta	tctagctcag	tttgctctac	aagacctgca	atagagcctt	gtgtgacata	12960
	gagataatat	ttgttgaagc	aattaaattt	gacttggaat	taactctgcc	atcattctat	13020
	aaggaaggat	tgaaaatcct	tctcaccctg	tgctgatata	gtacctttct	atacaaaaac	13080
	gtccttctcc	ctcttccctt	ggattgcata	aactatgtac	atgccttcct	caggggcact	13140
25	tttctaggac	agtgtcagcc	taaggatctt	tgtttgggtg	gcttttagaa	actcaggaag	13200
	acaggagcat	catatgccta	taggcagctg	gcttccaggt	cagtagtttt	gctctgaccc	13260
	taaaatcaga	ctcccatccc	aatgagtatc	tacaggggag	gaccgggcat	tctaagcagt	13320
	ttacgtgcca	attcaatttc	ttaacctatc	tcaaagatgg	agatcagtgt	gagtccaatc	13380
••	catgtttaaa	tggcggcagt	tgcaaggatg	acattaattc	ctatgaatgt	tggtgtccct	13440
30	ttggatttga	aggaaagaac	tgtgaattag	gtaagtaact	attttttgaa	tactcatggt	13500
	tcaaagtttc	cctctgaaac	aagttgaaac	tggaaaatgc	aatattggtg	tatcataatt	13560
			gatgcttata				
			attaaaatca				
25			cttattaatc			_	
35			gatgagaaaa				
			acttaaactt.		_		
			aaactgaacc			_	
			ttaagtacag				
40			gtacaaacca				
40			actatattga				
			ttctcaaagc				
			agtgtccctt				
			gaaacgcaca				
15			gatgcaatcc				
45	gactaaggat	aaagagtaga	aaattagcag	gactctatta	accgattaca	gcaatccacc	14400



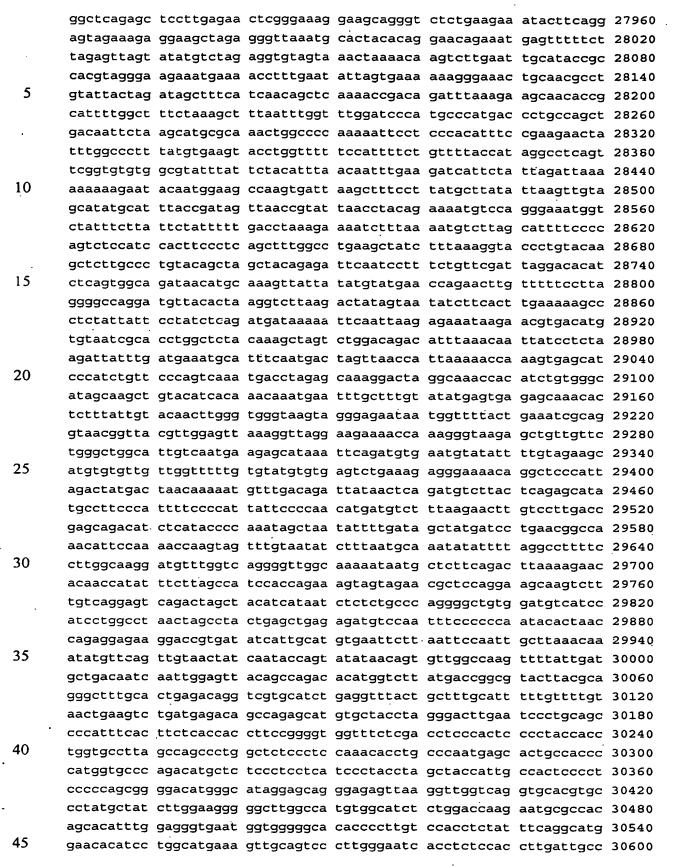


	aggttagtcc	acctgagtgg	gaagaagcgg	tgaaagaatt	tgtctgtcag	tatctgtctc	17160
	tcattggtta	gaagttcgac	ttatggggaa	ttaactccct	cacatttcct	agttggatag	17220
	cttgggtacc	agaggcatat	ggcatccatg	ctcagcatga	acagggaagc	ttcaaggcaa	17280
	aagacacata	gtgcagctat	gagccaaggc	aattcaagga	tacacccata	ggaggctggt	17340
5	tgacatccac	ccagagctaa	tcaccaccat	gctggaaaaa	gacacaggtg	aagctgagaa	17400
	gaatgaaggt	ggtgcatagg	aggtatctaa	tacagtcact	cattttcaaa	ctttccatgt	17460
	tatgattgca	ctgaccactg	aggatttcta	ttgaaagttt	tactgttgtc	aaacacgtac	17520
	acaaggggaa	aggtgtctta	cattgtttat	gttcctgtgc	tgctctagaa	acagaaatag	17580
	gctcaagagc	agagcctgtt	tttcttaatt	cagcaggtct	aagctaacaa	gtcctgaaac	17640
10	atggtacttc	ctgttattgg	tattgcatag	gagaaacaaa	gggaaagcac	agtaattaga	17700
	aaatacaaac	aagatggcag	gaataagcca	aaaatatcag	gaaacacaat	tattgtgaat	17760
	tgggattaaa	ctaatctatt	aataatgaca	actttcagct	tggagttaaa	aatttaattg	17820
•	tatactgtta	acgaaagtga	tacctaaaat	aaaattacac	tgggaggcca	aaatgaaggg	17880
	atgtgaaaag	aactatcagg	taaaaactaa	caaaaagaaa	ctagcaaagc	aatcttaata	17940
15	tcagacaaaa	tagaatccaa	gaggaaaatc	atttcaaaag	acaagagatt	ttttttatta	18000
	ataaggggaa	ttgcatagga	gagtaaagaa	aatgtgggcc	actggaatgc	ttagcactaa	18060
	tgacatattg	gtctttggtc	ttcagttacc	ttacaggacc	ctatttcatt	ctcttatgtt	18120
	tgatatgtaa	ccacctcagc	cagcttcaag	ttgctttttg	gccctaatgg	acttcctagc	18180
	actataattt	cttttttt	aaatgtttta	ttttaggttt	aggggtacat	gtgaaggttt	18240
20	gttacataga	taaacatgtg	tcacaggggt	ttgttgtaca	tattattaca	tgacgcagat	18300
	attcagctca	gtaccaaata	gtgatctttt	ctgctcctct	gcctcatccc	accctcctcc	18360
	ctcaagtaga	ctccagtatc	tgttgtttcc	ttctttgtgt	ttataagttc	ttaacactta	18420
	gctcccgctt	acaagtgaga	acctgcagta	tttgattttt	gttcctacgc	tagtttccta	18480
	aggatgatag	cctccagctc	cattcatatt	cccacaaaag	acataatctc	cttctttct	18540
25	atggctgcat	aatattccat	ggtatatatg	aaccacattt	tctttatcca	gtctgtcatt	18600
			ccatgtctgc				
			aacctgttaa				
			aattacactt				
20			ttgtgttcac				
30			agaaaaccca				
			gagtcaacat				
•			ttataatagc			_	
			ctacataaaa				
35			gaataggcaa				
33			tggattcaat				
			attcataagc			_	
			gttttgagga				
			gttacattta				
40			acactagagt			_	
. 40			gccagtgcag			_	
•			atattcatat				
			aaaaatttct				
•			atattaaaac			,	
45			acagagtgag		_	•	
47	actgaacccc	atacagtgat	ggtgggaatt	taagttcgta	caatcatttt	agaaaattgc	19800

	ttggcagtat	ctactagatc	tgaacatgtg	atccagtaat	tacactcata	attataaqcc	19860
		catgtttatg				_	
		ccaaaaactg					
		atagtaatac					
5 .		atggtttaat			_		
		cttctgttta					
		ggtaacctat					
		gataatgtgc					
		cttacactaa					
10		taaacctgta					
	tttaaaccaa	tcaattatag	tgctaccatc	atttttatgc	attattgaga	agtttattt	20460
	acctttcttt	ccactcttat	ttcaaggctc	caaaatttct	ctccccaacg	tatattgggg	20520
	gcaacatgaa	tgcccccaat	gtatatttga	cccatacatg	agtcagtagt	tccatgtact	20580
	ttttagaaat	gcatgttaaa	tgatgctgtt	actgtctatt	ttgcttcttt	tagatgtaac	20640
15	atgtaacatt	aagaatggca	gatgcgagca	gttttgtaaa	aatagtgctg	ataacaaggt	20700
•	ggtttgctcc	tgtactgagg	gatatcgact	tgcagaaaac	cagaagtcct	gtgaaccagc	20760
	aggtcataat	ctgaataaga	ttttttaaag	aaaatctgta	tctgaaactt	cagcatttta	20820
	acaaacctac	ataattttaa	ttcctacttg	aatctgcttc	cttttgaaat	catagaaaat	20880
	atcagtagct	tgaattagac	caattaattt	tctagattgc	atcatattt	aaatataact	20940
20 .	atgtaatcat	ctacaacctg	aattctttct	gtgtccaatt	tgtccaattt	ttttctctaa	21000
	catttatatc	acaaagcaat	taatttgtgt	gatttctgca	tatgtatttg	taattcatca	21060
,	agtcaaatca	atgtagtaat	actatatcat	aaaatataca	caaataattg	agtgataggc	21120
	ttctagtata	aggacggtaa	gtttgaagca	tgattctatc	tgggctggct	agtttactct	21180
	gagaaagtta	ttttttattg	ttgggtctta	agctgagttt	acacacttgg	tgtcagaatg	21240
25		tgaactgttt					
		atgtttccca					
		atctccactc					
•		ttttgaaaaa					
20		agcaaactag					
30		ttaaattgtg					
		agaagaaata					
		gcactagagg					
		gtctcacagc					
35		gcaagccagt		_			
33		gctagttatc					
		ggtttaagaa					
		attctggtaa					
		tgctgcagag					
40		ctgttcccaa					
70		aaggetteag					
		gctagtctac					
		cttaaaaacc tctatgctat					
		gaatatagcc					
45		tttcaatgat					
,,,	TETTETT	cccaacyac	aaayyeeeet	ccccigcage	carryrryat	Lycycacact	ZZ50U

	tatacataag	tattttgaac	taatttcctg	ttttctcaac	cacttgctgt	cttcatgata	22560
						tgaaggaaag	•
						attctggtag	
						ggaaaaatgg	
5	tgggcaggtc	ctggtgaata	tggctgtgat	aattatatta	gcaatctctt	tggctaatat	22800
	ttgaagccca	aataattgaa	tcacaatgat	ctctccccag	aaaatatata	aaatgcacct	22860
	tggaatctag	aaggcctttt	agtctgcaaa	agaaaccttc	ttaatcataa	gcagcagaag	22920
. '	tcccatttac	caaattggaa	agttaaagtt	acaaagcatc	aatcatcaga	cttccattca	22980
	gggatggcaa	ttgggagtaa	gactttttag	taaagaaact	aaacacaaag	tcattagact	23040
10	ctgtaaaagt	cttaccaaat	ttgattctgg	aacacctatt	ctatttccgt	aaagatgatg	23100
	aattcggagc	caaatgttct	tttcatgaag	gatttgaaaa	ctgtccatga	aaataacgca	23160
	atcaaccttt	tagcttgaga	ctctattcac	tgattagatt	tttttaaata	ctgatgggcc	23220
	tgcttctcag	aagtgacaag	gatgggcctc	aatctcaatt	tttgtaatac	atgttccatt	23280
	tgccaatgag	aaatatcagg	ttactaattt	ttcttctatt	tttctagtgc	catttccatg	23340
15	tggaagagtt	tctgtttcac	aaacttctaa	gctcacccgt	gctgaggctg	tttttcctga	23400
	tgtggactat	gtaaattcta	ctgaagctga	aaccattttg	gataacatca	ctcaaagcac	23460
	ccaatcattt	aatgacttca	ctcgggttgt	tggtggagaa	gatgccaaac	caggtcaatt	23520
	cccttggcag	gtactttata	ctgatggtgt	gtcaaaactg	gagctcagct	ggcaagacac	23580
	aggccaggtg	ggagactgag	gctattttac	tagacagacc	tattgggatg	tgagaagtat	23640
20 (	ttaggcaagt	ttcagcacta	accaatgtga	gaaggcctcc	agagatgagc	agttggtgaa	23700
	agagaggctc	aaaaccagct	accatacagg	tcaagaagaa	tttggcatta	aggaaacagc	23760
	atagcaggat	tccagacagg	caactggtca	acaacatgaa	ggtctggaag	aaaggtcgca	23820
	gtactcaggt	tcagggcact	acttcagctt	cagcccttgc	aaaaactggt	gagagttgga	23880
	aagtctttag	gctaagaaaa	attggattat	ttaaaagggg	taaagaaagg	gactcaagga	23940
2,5						gtcttgatct	
	accactatag	ttctcgtggt	agcatcagaa	tcacctggga	acgtagaaat	gcaaattctc	24060
						gcagtctgtg	
						aggccaggtg	
						attgcttaaa	
30	cccaggagtt	tgagaccagc	ctgggcaaca	cggcgaaacc	ccacctctaa	ttaaaaaaaa	24300
					_	aggaggctga	
	•					tcacaccact	
						agaaaaacac	
25						ataatgcact	
35					_	tgcaaacaac	
						tacatttgta	
						tctgtgagaa	
						tccttatata	
40						ttaaattaac	
. 40					-	tttaagcaca	
						atagaacact	
						acttaaaact	
•						tgaggttaga	
15.						gtgacaattt	
45	aagccagaat	caggtaatgt	cctcagtttt	aacagcattg	gaattttctg	ggactagctg	25200

	tgtatctatc	caggattctt	gagaatgcct	gccattttc	aacataatgg	atgtaaggta	25260
	ttacacatat	acctggggat	ggggaggtag	gtataattgc	acaagcattg	tggagaatgg	25320
•	tatcaaagag	tggcagaaca	tcacaatcaa	ggttttccct	ttettttacc	tttgcttttt	25380
	aaaaagacaa	tatttgctgg	acctgatctt	ataactcata	aatgggacac	tgtatgttcc	25440
5	tttttacctc	ctctgtttct	acttaattgc	accctatgag	gactgcttcc	cttacctacc	25500
	ataacccctt	ccttcactca	tccatatctt	tactcttctt	cacaactctg	taatattgac	25560
	cttctttatg	aacctttcct	ggaacaatcc	ctcttaatgc	aagcactgtt	attatgcctt	25620
	caatgtattt	aatatccatg	tatctattct	ctctaatttt	gtcattttgt	gttctcatgt	25680
	attttcattc	attatgtgtc	caacttccat	ggataacatg	gttacaacaa	aagatcctac	25740
10	tttatgacaa	ttatcttcct	tgggtttgtg	ggacatagaa	cagtgcacag	agtaggggat	25800
	ccaagaaccc	aggagaatat	attagctaag	aagataactt	ccgtttttaa	aagtccaaga	25860
	ttċaggagat	caaaaccatc	ctggctaaca	tagtgaaacc	ccgtctcttc	caaaaataca	25920
	aaaaattagc	ccggcgtggt	ggcaggcgcc	tatagtccca	gctacacggg	aggctgaggc	25980
	aggagaatgg	cgtgaaccgg	ggaggcggag	ctggcagtga	gccgagatcc	cgccactgca	26040
15	ctccagcctg	ggcgacagag	cgagactccg	tctcaaaaaa	aaaaaaaaa	aaaaagtcca	26100
	agattttaaa	aaaaaaaaa	aaaaggatgt	ctgctttgtg	agtttagcat	tgtctccttg	26160
	tcattccaga	aatgaaatgg	caaatacatt	taaatcagaa	ctaaaaaggg	gaacagggta	26220
	taaaggctca	atttagtcac	atcatttccg	tttctcaccc	accccctta	aaccagatgt	26280
	ttgccaatgc	attaacaatg	cagatgtttc	ctgaaagaaa	gtttagtaac	tcaagcagac	26340
20	accttattt	cttttcaagc	agaaaagact	atgagatggt	ggttgtggtt	gttccgggag	26400
	ggagaagata	taaatgatac	acattatttc	aaatcatttc	atgacctcac	tgcacactta	26460
	tagttattgt	acctgttgtc	tttttgctgt	caagcctagc	taagatcatt	tggaatgttc	26520
	aagatcactc	atacatgcat	gtgcacacat	acacatgcac	atatgttcac	tccctatttc	26580
	atccacatga	actaagatta	ctgatgtgta	cagattcaaa	gcacttttat	tcttttccaa	26640
25	aggcaagaag	ctgagctact	ttccagaata	gttgtgaaag	accctgtcat	acttctgcat	26700
	tgtttcctcc	acaccacctc	catccagttc	cttatgaatg	gttactggtt	ttcaaaaata	26760
	tgagataaat	tgagtgtata	aaagtcattt	ttagacaaaa	tgaaacagga	aatgaaagaa	26820
	accagaatct	ctcctcattt	gtggatgggc	cagctccacc	atgtcatggt	taatctgcag	26880
	ggaggaaata	ctagatttga	ttgcagatca	gactgcagca	aacctgctgt	gactaaggca	26940
30	tcaagagaaa	gcaagcaaca	gactggggct	tcagtggtga	aaacattata	tatctagctt	27000
	tgaaatatga	aatactgttt	agcagtgtca	cctagaaaag	agtgtttcaa	aatgctgatg	27060
	cttcataaga	acctttctct	tcagagttgg	tttcttttat	ctttcaaatt	agccagggtg	27120
	ggaaataaag	tgatcacttg	gtgaagaaat	ctcacaaaga	agaacataga	gagttcactt	27180
	tcatctggag	taatgaacag	attgaacaaa	ctagaaatgg	ttagtctgtt	aaagaaaagg	27240
35	tgtaggtgag	ctgtttgcaa	gagccacaag	ggaaagggga	agacaacttc	tttgtggact	27300
	taagggtgaa	agttgcaagc	aggcaagacg	attctgacct	ccattaagaa	agcccaaacc	27360
			gttacgcagg				
	aacaaatgtt	tgtcggaatt	gttgacttaa	agagctgttc	tgtcactggg	gacagcagcg	27480
			gagagggcat				
40	taagggactg	ctgggatcct	gtccagcttt	gagaccctac	agagccatgt	tcacctagca	27600
			ggtcatttct				
			caacgcagcc				
	agccccctat	ggttccggaa	cgcgctgccg	gccccatcct	gtttgctacc	tcctaaagcc	27780
4.5			ccagcttcta				
45	ggaacggcgt	gaggccaatg	gaaggaggta	cttcagtttc	cctccagatg	cccagcgatg	27900





	acagtaggcc	agtgacaagg	gaagattgac	acatcatccc	tgctggggcc	cagtgtcctg	30660
	tggctggcag	gcaggggatc	ctaaggacat	gtgggtctta	aattgtaggg	tgcacttcct	30720
	gggcaccttt	gagggtctgc	actgccccag	caaatatccc	catgctagaa	ggagcaaaat	30780
	attaaatggc	aaattttaaa	aatgtaacaa	gatgggttgc	aaaagagact	acagaggaaa	30840
5	gcaaaagttt	tgtattttag	tatcttccat	ggcacttttc	ttcctagctt	ttgaacaagg	30900
	ggccccacat	ttttatttct	cactgagccc	cacaaagtat	gtagccattc	ctgcccggag	30960
	tgaggacttt	taaaacataa	agattatcaa	gtcttggaaa	ttctgattca	gtagatatat	31020
	aacaggtctc	aaacttaatt	atgtaaagaa	tattctggag	agcttccttt	tacccagtcc	31080
	cacccaccaa	atattctgat	aaattaagct	tcgattagcc	cccagatctg	cattttataa	31140
10	ggatccccag	atgattctac	tgcaattggt	ccacagacca	tgcctggacc	gaatttgggt	31200
	gcttaggagc	acaaattctg	gagccgggca	gacttgagtt	tgcttcctag	ctttaccaac	31260
	tgatctcagg	ggagttaatg	tttacctcta	aactttagct	catgcatcta	taaataaata	31320
	tattaatatc	atgtcataag	gatattatgt	tgtattaaat	gtctttaaaa	caccacaatg	31380
	attagcccaa	agtaaacact	caataaatgt	tcaaaaattt	aggaaaattg	ttaagactgg	31440
15	gttgtatgca	cactggtgtt	tattatatta	tgtagttttt	tctgtatttt	tacaacattt	31500
	cagaattaaa	agcaacagct	agaaaaagag	ggaaatggcc	gggtgcagtc	gtcacgcctg	31560
	taatcccagc	actttgggag	gccaaggcgg	gcggatcacg	aggtcgagag	atcgagacca	31620
	tcctggccaa	catggtgaaa	ccccatctct	actaaaaata	caaaaattaa	ctgggcatgg	31680
	tggcatgcgc	ctgtagtccc	aggagaattg	cttgaacctg	ggaggcggag	gttgcagtga	31740
20	gccaagatct	caccactgct	ctccagcctg	gtgacagggc	aagactccgt	caaaaaaaaa	31800
	aaagagaggg	agagccagag	tatgaaaaag	gaagtcagag	ccctttaatg	agtcagcttt	31860
	gtaggtctcc	aggtaggagg	ctagtgcttc	agtgtctagg	acatagtagg	tgttcagtaa	31920
				aaggaccatc			
				ggttcatgcc			
25				tcctgaaaag			
				cttcctataa			-
				atctagacat			
				agttgagata			
20				tgtccacttt			
30				actatccagc			
				ctgctattcc			
				gcttcacaat			
•				tttctctttt			
26				tagtaactat			
35				ccatgaagca.			
				aaatcaccaa		_	
				tgtgtgttgc			
				aattcatctg			
40				aatattttgc			
40				tatgcagtaa			
				ttctgtggag			
				actggtgtta			
				accactagct			
45				gttgaataaa			
7.7	LCALARETEC	agaacccacg	regeacegre	ctccaagcat	ccatagttct	tttgatatac	3300

	ccctattatc	actcatttca	gtgaggtaca	attagttctt	gatgtagcca	tttccatacc	.33360
	agaaggcctt	cccaaaaatc	agtgtcatgt	caccgatcct	tttatctctg	gtgcttggca	33420
	caacctgtag	caggtcctca	gaaaacaaac	atttgaatta	atggccaaat	gagtttgtgc	33480
	tcaaaaaagg	ggtgaggata	cttgaaattt	ggaaaatcta	ggataattca	tgactagtgg	33540
5	attcattatc	accaatgaaa	ggcttataac	agcatgagtg	aacagaacca	tctctatgat	33600
	agtcctgaat	ggctttttgg	tctgaaaaat	atgcattggc	tctcattaca	tttaaccaaa	33660
	attatcacaa	tataagaatg	agatctttaa	cattgccaat	taggtcagtg	gtcccaagta	33720
•	gtcacttaga	aaatctgtgt	atgtgaaata	ctgtttgtga	cttaaaatga	aatttattt	33780
	taataggtga	acataatatt	gaggagacag	aacatacaga	gcaaaagcga	aatgtgattc	33840
10	gaattattcc	tcaccacaac	tacaatgcag	ctattaataa	gtacaaccat	gacattgccc	33900
	ttctggaact	ggacgaaccc	ttagtgctaa	acagctacgt	tacacctatt	tgcattgctg	33960
	acaaggaata	cacgaacatc	ttcctcaaat	ttggatctgg	ctatgtaagt	ggctggggaa	34020
	gagtcttcca	caaagggaga	tcagctttag	ttcttcagta	ccttagagtt	ccacttgttg	34080
	accgagccac	atgtcttcga	tctacaaagt	tcaccatcta	taacaacatg	ttctgtgctg	34140
15	gcttccatga	aggaggtaga	gattcatgtc	aaggagatag	tgggggaccc	catgttactg	34200
	aagtggaagg	gaccagtttc	ttaactggaa	ttattagctg	gggtgaagag	tgtgcaatga	34260
	aaggcaaata	tggaatatat	accaaggtat	cccggtatgt	caactggatt	aaggaaaaaa	34320
	caaagctcac	ttaatgaaag	atggatttcc	aaggttaatt	cattggaatt	gaaaattaac	34380
						tatacattct	
20						gcaaattgat	
	tagaaaatgg	aaccactaga	ggaatataat	gtgttaggaa	attacagtca	tttctaaggg	34560
						atactatggt	
						catcttcccg	
						cagtacagga	
25						cacaggagta	
						ttcctcaatc	
						tccctctctc	
						tatacttctg	
20 :						cttgcttttc	
30						tttcagagag	
	•					ataaatatat	
•						atggaagcaa	
						ttcacgaagg	
35						attgtactct	
33						tgcaatctac	
						tcagagccaa	
						ttaactagca	
						aatcagccaa	
40						agttaatcct	
						tttgtgaatt	
						tgtgaatcag	
						ctgccaagct	
						agttccttta	
45						tatttttaaa	
	tttcttggct	aggeceageg	guicacycol	acaaccccag	cacticityga	ggccaaggcg	20000

	ggcggatcac	ctgaggttag	gagtttcagg	ccagcctggc	caacatggtg	aaaccctgtc	36060
	tctactaaaa	atagaacaat	tagctgggct	tggtaatgtg	cacctataat	cccagctact	36120
	ggggaggctg	aggcaggaga	atcacttgag	cctggggagc	agggggtgcg	ggaggttgca	36180
	gtgagacaag	atcgcaccag	tgcactcccc	atcctgggtg	acagagtgag	actctgtctc	36240
5	aaagaaaata	aataaataaa	tacatttctt	gaggcgtttc	ttgttaaatc	attcatggag	36300
	aggcatccca	aacaccacat	tcaacaaaac	actctgaaaa	atgttttcaa	atgcaatata	36360
	acacagcaga	gatttgatgc	tctgttatcc	agttttcata	taaggctgtg	tgagctgtgt	36420
	cccagagagg	acagtggtct	gaatccacct	gagacagaat	tgggtctaac	taactgtgag	36480
	tatggccttc	aataagtcac	tctccatttg	ggaatttgat	ttctccactt	gtataatgag	36540
10	agtatttgac	aggatgctct	cccaaatccc	ttgcaatttt	gttagtctgt	gatttcatgt	36600
	ttttatttt	attccttcat	ccaacaaata	gtcaaggagt	aattgctgtg	tgccaaatac	36660
	caacagtatt	cattaaattg	taattcagat	tttatatata	tataaataat	gtataatgtg	36720
	tataaattgc	tttgtgagtg	cctactacac	tgctagacag	tagttgctca	ataacttgtt	36780
	agctgaatca	gaatccatgt	ttatcccaga	gtagcaatta	gtcttgcatc	gagtatcgtg	36840
15	aaagaaggcc	acacttaaat	aagaataatg	cctggggttt	aggttttatg	aaaaaatgaa	36900
	aggaaattag.	ttctgctttt	gttgactaaa	ggaagggaag	agagaagaga	cactataatt	36960
	gtctgcctca	gatttaagga	ggaggctaat	tcatgcatta	aacacgttac	ttcaaatttg	37020
	aatgaccaaa	ggtctgtagc	ctcagcactt	caaaattggt	aaaagtaaga	cactctggcc	37080
	ttgtttccat	agagaccacc	ccttacaaag	gcaccaatgg	gaaactggcc	tcaggactcc	37140
20	tgttattggt	cttctctgtg	gcagagaaag	gagctcttgg	acccataaat	ctctgagcca	37200
	cagttctttt	tgccatgggc	tcaaaaatga	ttgaattcat	catgagccac	ctgtggcata	37260
	ttgccacact	aaacatgtgg	ggcctttaag	ctcactaaga	gccaatgtct	tcagagccag	37320
	ccctggcttg	attctaccta	gggcatttgc	agttgccata	taagaatcat	tagtgctttc	37380
	aaaattactg	tagatacttt	gcctaaatag	actaaaacat	gctgccgtca	tattggaagt	37440
25	gacagattaa	aatagaactc	ttgccaagtg	aaggaaagtg	tgctaatata	atgcagtcat	37500
	tttaacttgc	tgtttaagtg	tgattgtttt	tagttctttt	gaatattatt	tgttttatac	37560
	tgacaggaac	gaagtactgt	ccaattttct	ctgccaagga	aaaaagaaaa	ggtgttcttc	37620
	cttacttacc	tgaaccaaaa	cagaccagtt	tacaaaattg	cctaattata	attgctaaac	37680
	aagttccgaa	tgcttacagt	ctaatccaag	aatgtcagag	ctgcaagggc	ccttaaacac	37740
30	catccaatcc	actccactca	tttagcagat	gaagagattg	agggcaacat	aaggccaggc	37800
•	ccaagataac	acaatgacag	ccaggactag	agctcaagtc	tcccaccctg	cactttgaaa	37860
	gaataatgct	ttcaactgga	gtacattaac	tctactgtct	atatttttag	ggcagctggg	37920
	gcattctgca	ttggtggcaa	tcctctcaac	aaccctggga	ctgaaaactg	cctggaattc	37980
_	ttactaacaa	ttctctaatt	gaccaaaagg	tgacgaaatc	aaggagacca	ataaggtagc	38040
35	cttggaaagc	aagagtggc				•	38059

<210> 5

<211> 29

<212> PRT

40 <213> Homo sapiens

<400> 5

Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
1 5 10 15

5 Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr
20 25

<210> 6

<211> 55 ·

10 <212> PRT

<213> Homo sapiens

<400> 6

Val Phe Leu Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys

15 1 5 10 15

Arg Tyr Asn Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu
20 25 30

20 Arg Glu Cys Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val
35 40 45

Phe Glu Asn Thr Glu Arg Thr 50 55

25

<210> 7

<211> 8

<212> PRT

<213> Homo sapiens

30

<400> 7

Thr Glu Phe Trp Lys Gln Tyr Val

35 <210> 8

<211> 38

<212> PRT

<213> Homo sapiens

<400> 8

Asp Gly Asp Gln Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys

1 5 10 15

5 Lys Asp Asp Ile Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu
20 25 30

Gly Lys Asn Cys Glu Leu

. 35

10

<210> 9

<211> 43

<212> PRT

<213> Homo sapiens

15

<400> 9

Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe Cys Lys

1 5 10 15

20 Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly Tyr Arg 20 25 30

Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala 35 40

25

<210> 10

<211> 68

<212> PRT

<213> Homo sapiens

30

<400> 10

Val Pro Phe Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu

1 5 10 15

Thr Arg Ala Glu Ala Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr 20 25 30

Glu Ala Glu Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe
35 40 45

40

Asn Asp Phe Thr Arg Val Val Gly Glu Asp Ala Lys Pro Gly Gln 50 55 60

Phe Pro Trp Gln

45 65

<210> 11 <211> 38 <212> PRT <213> Homo sapiens <400> 11 Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly Val Lys Ile Thr Val Val Ala <210> 12 <211> 182 <212> PRT <213> Homo sapiens <400> 12 Gly Glu His Asn Ile Glu Glu Thr Glu His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His Glu Gly Gly Arg Asp Ser Cys 

- 19 -



Gln Gly Asp Ser Gly Gly Pro His Val Thr Glu Val Glu Gly Thr Ser 130 135 140

Phe Leu Thr Gly Ile Ile Ser Trp Gly Glu Glu Cys Ala Met Lys Gly

145 150 155 160

Lys Tyr Gly Ile Tyr Thr Lys Val Ser Arg Tyr Val Asn Trp Ile Lys
165 170 175

10 Glu Lys Thr Lys Leu Thr 180

<210> 13

<211> 2410

15 <212> DNA

20

25

30

35

40

45

<213> Mus musculus

#### <400> 13

gaattccgtg gatgtgcttt aaaaccacac ctaacgtttg agcacaagtc tcacgaactg 60 gegetacaac tteateagaa acgaagtete caaatetgte caacgeaaaa acacaaagga 120 gtctaatgac taagtcttcc aaccacaact gtctgctgcg cccggaaaac aagccggggc 180 tetggggace eggggeteag geegeetege teeggeetag eeeegeeace ttagttgtgt 240 cateceeegg geatgetgag cateceeeeg eggeteegge acagaegeee ggaeeteagg 300 tetetgeete egegegggg ceeggeeetg tggeeggagg gageggeegg atggagegga 360 ggatgaaagg cggatacttg gaccagcgag tgccctacac cttctgcagc aaatctcccg 420 gaaatgggag cttgggcgaa gcgctgatgg tcccgcaggg aaagctcatg gacccgggct 480 ccctgccgcc ttccgactca gaagatctct tccaggacct cagtcacttc caagagacgt 540 ggctcgcaga agctcaggta ccggacagtg atgagcagtt tgttcctgat ttccattcag 600 aaaacttagc tttccatagc cccaccacca ggatcaagaa ggaaccccag agtccccgca 660 cagaccccgc cctgtcctgc agcaggaagc caccactccc ctaccaccat ggagagcagt 720 geetttacte cagacaaate geeatcaagt ecceegetee eggtgeeeet ggacagtege 780 ccctgcagcc cttttccagg gcagaacagc agcagagcct cctgagagcc tccagctctt 840 cccagtecca ecctggecae gggtacettg gtgageaeag etcegtette cageageeeg 900 tggacatgtg ccactcettc acateteetc agggaggggg ccgggaacet eteccagece 960 cctatcaaca ccaactgtcg gagcctgcc caccctaccc ccagcagaac ttcaagcagg 1020 agtaccatga ccccctgtac gaacaggctg gccagcccgc ttcaagccag ggtggggtca 1080 gtgggcacag gtacccaggg gcgggggtgg tgatcaaaca ggagcgcaca gacttcgcct 1140 acgactcaga tgtccctgga tgtgcatcaa tgtacctcca cccagagggc ttctctggac 1200 cetetecagg tgatggagtg atgggttatg getatgaaaa atccettega ceatteceag 1260 atgatgtctg cattgtccct aaaaaatttg aaggagacat caagcaggaa gggattggag 1320 ctttccggga ggggccaccc taccagcgcc ggggtgcctt acaactgtgg cagtttctgg 1380 tggccctgct ggatgaccca acaaatgctc atttcattgc ttggacaggc cggggaatgg 1440 agtttaaact aattgaacct gaagaggttg ccaggctctg gggtatccag aagaaccggc 1500 cagecatgaa ttatgacaag etgageeget egetgegata etattatgag aaaggeatea 1560 tgcagaaggt ggctggcgaa cgctacgtgt acaagtttgt gtgcgagccg gaggccctgt 1620

```
tetetetgge etteceagat aateaaegte eagetetgaa ggetgagttt gaeeggeeag 1680
      tcagtgagga ggacacagtc cetttgteec acttggatga gagteetgee taceteecag 1740
      aactcactgg cccgctccg cccttcggcc acagaggtgg atattcttac taggcaccag 1800
      tggcttcccc ttgacatggt ggggttgctc agtgtatata tcaactgatt tggtattggt 1860
 5
      gaaggccctc tttctgatgc ctgtagaagt ctctggggtc agagctccac tatcccatct 1920
      gatactectg gecagaetea getgetaace agagtetgeg ggaaagaeag tggaggeagg 1980
      ccaaatctaa aggcagtagc tgaagttcgc tgtggctcac ctgtaccttc agttcagctt 2040
      ggcctctgcc taggtcttgc tcagaggcca agttcctcac ccccaccaca gagatccagt 2100
      gttctattct ggggacatac agggacttcc cttgtttatt atggcaacag ggccaagggg 2160
10
      attotoagaa caccotgtgt otococtoto coaaccocco atgggagaca aagttotgco 2220
      tggcttctgc cctgaacagg ggggtcctgt gttcttggtg ctgtgctctg ggaggcagga 2280
      gcatgtgggc ggcagctggg ggggggtgtg gaagtagaga tggctctctg ccctaggcct 2340
      acccaggeet aattecaeet ttgeetetta tgeeagaeet taataaagee tetgettete 2400
      cccggaattc
                                                                         2410
15
      <210> 14
      <211> 32
      <212> DNA
      <213> Artificial Sequence
20
      <220>
      <223> Description of Artificial Sequence: Synthetic
      <400> 14
25
      taacaggatc cggcctctca ctaactaatc ac
                                                                         32
      <210> 15
      <211> 31
      <212> DNA
30
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
35
      <400> 15
      caactggatc caagattcaa gatagaagga t
                                                                         31
      <210> 16
      <211> 29
40
      <212> DNA
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
45
```

	WO 00/75279	PCT/US00/15728
	<400> 16	
	ggacaggate ecceaaactt tteaggeae	29
	<210> 17	
5	<211> 29	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
10	<223> Description of Artificial Sequence: Synthetic	
	<400> 17	
	caagcatgca tctagtgtta gtggaagag	29
15	<210> 18	
	<211> 29	
	<212> DNA	
	<213> Artificial Sequence	
20	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 18	
	caagcatgca aatattaact caaaatgga	29
25		
	<210> 19	
	<211> 28	
	<212> DNA	
	<213> Artificial Sequence	
30		
	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 19	
35	caagcatgct gttgtttttt gttttaac	28
	<210> 20	
	<211> 29	
	<212> DNA	
40	<213> Artificial Sequence	
	<220>	
	<223> Description of Artificial Sequence: Synthetic	

	W O 00/73279		· PC	1/0500/15/
	<400> 20			
	caagcatgca gccattcagt cgaggaagg			29
	<210> 21		•	
5	<211> 27			
•	<212> DNA			
	<213> Artificial Sequence	•	•	
	<220>			
10	<223> Description of Artificial Sequence:	Synthetic		
	<400> 21			
	caagcatgcg atcccttcct tatacct			27
15	<210> 22			
	<211> 27			
	<212> DNA		•	
	<213> Artificial Sequence			
20	<220>			
	<223> Description of Artificial Sequence:	Synthetic		
	<400> 22			
	taagcttaac ctttgctagc agattgt			27
25	•			
	<210> 23		٠	
	<211> 29			
•	<212> DNA `			
20	<213> Artificial Sequence			
30				
	<220>			
	<223> Description of Artificial Sequence:	Synthetic		
	<400> 23			
35	cttggtacca gccattcagt cgaggaagg			29
	<210> 24			
	<211> 29			
	<212> DNA			
40	<213> Artificial Sequence			
	· <220>			
	<223> Description of Artificial Sequence:	Synthetic	•	

	WO 00/73279	- FC1/0500/15/2
	<400> 24	
	cttggtacca tatgaatcct ttcatagat	29
	222	
5	<210> 25	
5	<211> 23	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
10	<223> Description of Artificial Sequence: Synthetic	
	<400> 25	
	ctgtgggaac acacagattt tgg	23
15		
13	<210> 26 <211> 19	
	<211> 19 <212> DNA	•
	<213> Artificial Sequence	
	(213) Aftifficial Sequence	
20	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 26	
		19
25	ggaataattc gaatcacat	19
23	<210> 27	
	<211> 20	
	<212> DNA	
	<213> Artificial Sequence	
30		
	<220>	,
	<223> Description of Artificial Sequence: Synthetic	
	<400> 27	
35	ccaatctgct cacacaggat	20
	<210> 28	
	<211> 20	
	<211> 20 <212> DNA	
40	<213> Artificial Sequence	
10	2217 Intilitat bequence	
	<220>	·
	<223> Description of Artificial Sequence: Synthetic	

WO 00/75279

PCT/US00/15728

### WO 00/75279 PCT/US00/15728 <400> 28 ccttgaggct gtccaagtga . 20 <210> 29 5. <211> 20 <212> DNA <213> Artificial Sequence <220> 10 <223> Description of Artificial Sequence: Synthetic <400> 29 atggttaact gacttacgaa 20 15 <210> 30 <211> 26 <212> DNA <213> Artificial Sequence 20 <220> <223> Description of Artificial Sequence: Synthetic <400> 30 gctccatttt gagttaatat ttgtgt 26 25 <210> 31 <211> 22 <212> DNA <213> Artificial Sequence 30 <220> <223> Description of Artificial Sequence: Synthetic <400> 31 35 ttcagtcgag gaaggatagg gt 22 <210> 32 <211> 22 <212> DNA 40 <213> Artificial Sequence <220> <223> Description of Artificial Sequence: Synthetic

	WU 00/752		_	PC1/USUU/15/2
	<400> 32			
	ttcagtcggt	tġgtgatagg gt		22
	<210> 33			
5	<211> 11			
	<212> DNA			
	<213> Homo	sapiens		
10	<400> 33			
10	tcgaggaạgg	a		11
	<210> 34			
	<210> 34 <211> 15	÷	•	
	<211> 13			
15	<213> Homo	saniens		
	1220			
	<400> 34			
	agtcgaggaa	ggata		15
20	<210> 35	•		
	<211> 19		•	
	<212> DNA			
	<213> Homo	sapiens		
25	_			
25	<400> 35			
	tcagtcgagg	aaggatagg		19
	<210> 36			
	<211> 23			
30	<212> DNA		•	
	<213> Homo	sapiens	•	
	<400> 36			
	attcagtcga	ggaaggatag ggt		23
35				
	<210> 37			
	<211> 27			
	<212> DNA			
40 :	<213> Homo	sapiens		
40				
	<400> 37			
	ccattcagtc	gaggaaggat agggtgg		27

## PCT/US00/15728 <210> 38 <211> 30 <212> DNA <213> Homo sapiens 5 <400> 38 gccattcagt cgaggaagga tagggtggta 30 <210> 39 10 <211> 7 <212> DNA <213> Artificial Sequence <220> 15 <223> Description of Artificial Sequence: Synthetic <400> 39 gaggatg 7 20 <210> 40 <211> 7 <212> DNA <213> Artificial Sequence 25 <220> <223> Description of Artificial Sequence: Synthetic <400> 40 caggaag ' 30 <210> 41 <211> 7 <212> DNA <213> Artificial Sequence 35 <220> <223> Description of Artificial Sequence: Synthetic <400> 41 40 caggatg <210> 42 <211> 12222 <212> DNA 45 <213> Homo sapiens

WO 00/75279

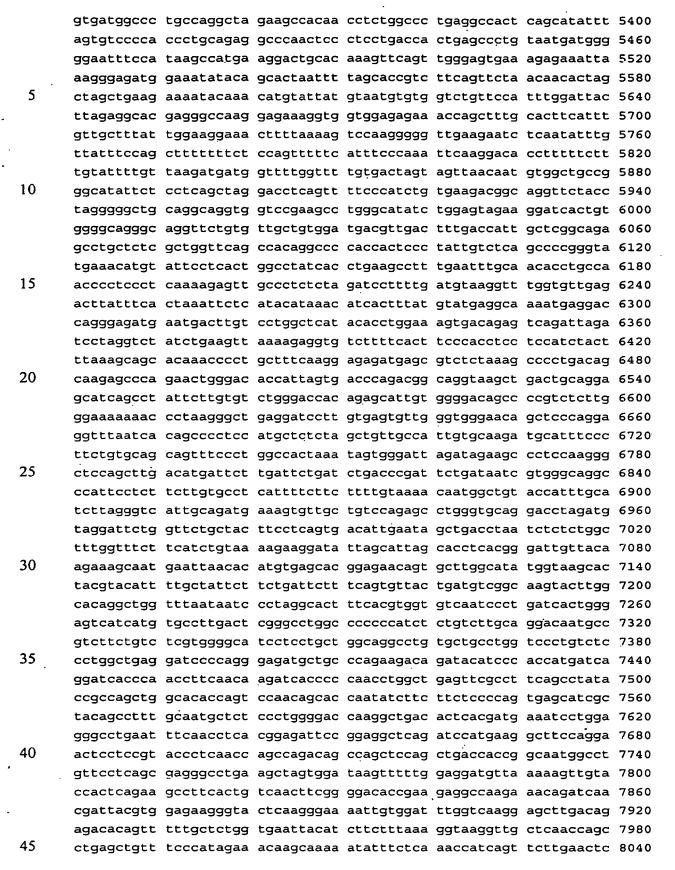
<400> 42

•	gaattccagg	ttggaggggc	ggcaacctcc	tgccagcctt	caggccactc	tcctgtgcct	60
•	gccagaagag	acagagcttg	aggagagctt	gaggagagca	ggaaaggtgg	aacattgctg	120
	ctgctgctca	ctcagttcca	caggtgggag	gaacagcagg	gcttagagtg	ggggtcattg	180
5	tgcagatggg	aaaacaaagg	cccagagagg	ggaagaaatg	cctaggagct	accgagggca	240
	ggcgacctca	accacagccc	agtgctggag	ctgtgagtgg	atgtagagca	gcggaatatc	300
	cattcagcca	gctcagggga	aggacagggg	ccctgaagcc	aggggatgga	gctgcaggga	360
	agggagctca	gagagaaggg	gagggagtc	tgagctcagt	ttcccgctgc	ctgaaaggag.	420
•	ggtggtacct	actcccttca	cagggtaact	gaatgagaga	ctgcctggag	gaaagctctt	480
10	caagtgtggc	ccaccccacc	ccagtgacac	cagcccctga	cacgggggag	ggagggcagc	540
•	atcaggaggg	gctttctggg	cacacccagt	acccgtctct	gagctttcct	tgaactgttg	600
	cattttaatc	ctcacagcag	ctcaacaagg	tacataccgt	caccatcccc	attttacaga	660
	tagggaaatt	gaggctcgga	gcggttaaac	aactcacctg	aggcctcaca	gccagtaagt	720
	gggttccctg	gtctgaatgt	gtgtgctgga	ggatcctgtg	ggtcactcgc	ctggtagagc	780
. 15	cccaaggtgg	aggcataaat	gggactggtg	aatgacagaa	ggggcaaaaa	tgcactcatc	840
	cattcactct	gcaagtatct	acggcacgta	cgccagctcc	caagcaggtt	tgcgggttgc	900
	acagcggagc	gatgcaatct	gatttaggct	tttaaaggat	tgcaatcaag	tgggacccac	960
	tagcctcaac	cctgtacctc	ccctcccctc	cacccccagc	agtctccaaa	ggcctccaac	1020
	aaccccagag	tgggggccat	gtatccaaag	aaactccaag	ctgtatacgg	atcacactgg	1080
20	ttttccagga	gcaaaaacag	aaacagcctg	aggctggtca	aaattgaacc	tcctcctgct	1140
	ctgagcagcc	tagggggcag	actaagcaga	gggctgtgca	gacccacata	aagagcctac	1200
	tgtgtgccag	gcacttcacc	cgaggcactt	cacaagcatg	cttgggaatg	aaacttccaa	1260
	ctctttggga	tgcaggtgaa	acagttcctg	gttcagagag	gtgaagcggc	ctgcctgagg	1320
	cagcacagct	cttctttaca	gatgtgcttc	cccacctcta	ccctgtctca	cggcccccca	1380
25	tgccagcctg	acggttgtgt	ctgcctcagt	catgctccat	ttttccatcg	ggaccatcaa	1440
	gagggtgttt	gtgtctaagg	ctgactgggt	aactttggat	gagcggtctc	tccgctccga	1500
	gcctgtttcc	tcatctgtca	aacgggetet	aacccactct	gatctcccag	ggcggcagta	1560
	agtcttcagc	atcaggcatt	ttggggtgac	tcagtaaatg	gtagatcttg	ctaccagtgg	1620
	aacagccact	aaggattctg	cagtgagagc	agagggccag	ctaagtggta	ctctcccaga	1680
30	gactgtctga	ctcacgccac	cccctccacc	ttggacacag	gacgctgtgg	tttctgagcc	1740
	aggtacaatg	actcctttcg	gtaagtgcag	tggaagctgt	acactgccca	ggcaaagcgt	1800
•	ccgggcagcg	taggcgggcg	actcagatcc	cagccagtgg	acttagcccc	tgtttgctcc	1860
	tccgataact	ggggtgacct	tggttaatat	tcaccagcag	cctccccgt	tgcccctctg	1920
	gatccactgc	ttaaatacgg	acgaggacag	ggccctgtct	cctcagcttc	aggcaccacc	1980
35	actgacctgg	gacagtgaat	cgtaagtatg	cctttcactg	cgaggggttc	tggagaggct	2040
	tccgagctcc	ccatggccca	ggcaggcagc	aggtctgggg	caggaggggg	gttgtggagt	2100
	gggtatccgc	ctgctgaggt	gcagggcaga	tggagaggct	gcagctgagc	tcctattttc	2160
	ataataacag	cagccatgag	ggttgtgtcc	tgtttcccag	tcctgcccgg	tccccctcg	2220
	gtacctcctg	gtggatacac	tggttcctgt	aagcagaagt	ggatgagggt	gtctaggtct	2280
40	gcagtcctgg	caccccagga	tgggggacac	cagccaagat	acagcaacag	caacaaagcg	2340
•	cagccatttc	tttctgtttg	cacageteet	ctgtctgtcg	ggggctcctg	tctgttgtct	2400
						cttctataga	
•			•		•	ctccatcccc	
4.5						atgggcctcc	
45	ccatgcaccg	tggacctctg	ggccctcctg	tcccagtgga	ggacaggaag	ctgtgagggg	2640



	cactgtcacc	cagggctcaa	gctggcattc	ctgaataatc	gctctgcacc	aggccacggc	2700
	taagctcagt	gcgtgattaa	gcctcataac	cctccaaggc	agttactagt	gtgattccca	2760
	ttttacagat	gaggaagatg	gggacagaga	ggtgaataac.	tggccccaaa	tcacacacca	2820
	tccataattc	gggctcaggc	acctggctcc	agtccccaaa	ctcttgaacc	tggccctagt	2880
5	gtcactgttt	ctcttgggtc	tcaggcgctg	gatggggaac	aggaaacctg	ggctgaactt	2940
	gaggcctctc	tgatgctcgg	tgacttcaga	cagttgctca	acctctctgt	tctcttgggc	3000
	aaaacatgat	aacctttgac	ttctgtcccc	tcccctcacc	ccacccgacc	ttgatctctg	3060
	aagtgttgga	aggatttaat	ttttcctgca	ctgagttttg	gagacaggtc	aaaaagatga	3120
	ccaaggccaa	ggtggccagt	ttcctataga	acgcctctaa	aagacctgca	gcaatagcag	3180
10	caagaactgg	tattctcgag	aacttgctgc	gcagcaggca	cttcttggca	ttttatgtgt	3240
	atttaatttc	acaatagctc	tatgacaaag	tccacctttc	tcatctccag	gaaactgagg	3300
	ttcagagagg	ttaagtaact	tgtccaaggt	cacacagcta	atagcaagtt	gacgtggagc	3360
	aatctggcct	cagagccttt	aattttagcc	acagactgat	gctcccctct	tcatttagcc	3420
	aggctgcctc	tgaagttttc	tgattcaaga	cttctggctt	cagctttgta	cacagagatg	3480
15	attcaatgtc	aggttttgga	gcgaaatctg	tttaatccca	gacaaaacat	ttaggattac	3540
	atctcagttt	tgtaagcaag	tagctctgtg	atttttagtg	agttatttaa	tgctctttgg	3600
	ggctcaattt	ttctatctat	aaaatagggc	taataatttg	caccttatag	ggtaagcttt	3660
	gaggacagat	tagatgatac	ggtgcctgta	aaacaccagg	tgttagtaag	tgtggcaatg	3720
	atggtgacgc	tgaggctgtg	tttgcttagc	atagggttag	gcagctggca	ggcagtaaac	3780
20	agttggataa	tttaatggaa	aatttgccaa	actcagatgc	tgttcactgc	tgagcaggag	3840
	ccccttcctg	ctgaaatggt	cctggggagt	gcagcaggct	ctccgggaag	aaatctacca	3900
	tctctcgggc	aggagctcaa	cctgtgtgca	ggtacaggga	gggcttcctc	acctggtgcc	3960
	cactcatgca	ttacgtcagt	tattcctcat	ccctgtccaa	aggattcttt	tctccattgt	4020
	acagctatga	agctagtgct	caaagaagtg	aagtcattta	ccccaggccc	cctgccagta	4080
25	agtgacaggg	cctggtcaca	cttgggttta	tttattgccc	agttcaacag	gttgtttgac	4140
	cataggcgag	attctcttcc	ctgcaccctg	ccgggttgct	cttggtccct	tattttatgc	4200
	tcctgggtag	aaatggtgcg	agattaggca	gggagtggac	getteeetgt	ccctggcccc	4260
	gcaaagagtg	ctcccacctg	ccccgatccc	agaaatgtca	ccatgaagcc	ttcattcttt	4320
••			gtgtccgtac				
30			ccaggcacta				
			aggttttgga				
			tgcttgagac				
			gctgtctcag			_	
2.5			gacttgctgt				
35			gaaaggagga				
		•	tgtgctcctc				
	•		ttgagcccag				
			tccatctgtt				
40			caggggctct				
40			caggagttca				
			aattagccag			•	
			agaatttctt	•			
			caccctggcg				
15			acagaggcc				
45	cacaactggg	gagagtgagg	ccccaggacc	agaggattct	tgctaaaggc	caagtggata	5340





	tccttggcaa	tgcattatgg	gccatagcaa	tgcttttcag	cgtggattct	tcagttttct	8100
	acacacaac	actaaaatgt	tttccatcat	tgagtaattt	gaggaaataa	tagattaaac	8160
	tgtcaaaact	actgacgctc	tgcagaactt	ttcagagcct	ttaatgtcct	tgtgtatact	8220
	gtatatgtag	aatatataat	gcttagaact	atagaacaaa	ttgtaataca	ctgcataaag	8280
5	ggatagtttc	atggaacata	ctttacacga	ctctagtgtc	ccagaatcag	tatcagtttt	8340
	gcaatctgaa	agacctgggt	tcaaatcctg	cctctaacac	aattagcttt	tgacaaaaac	8400
	aatgcattct	acctctttga	ggtgctaatt	tctcatctta	gcatggacaa	aataccattc	8460
	ttgctgtcag	gttttttag	gattaaacaa	atgacaaaga	ctgtggggat	ggtgtgtggc	8520
•	atacagcagg	tgatggactc	ttctgtatct	caggctgcct	tcctgcccct	gaggggttaa	8580
10	aatgccaggg	tcctgggggc	cccagggcat	tctaagccag	ctcccactgt	cccaggaaaa	8640
	cagcataggg	gaggggaggt	gggaggcaag	gccaggggct	gcttcctcca	ctctgaggct	8700
	cccttgctct	tgaggcaaag	gagggcagtg	gaggcaagcc	aggctgcagt	cagcacagct	8760
	aaagtcctgg	ctctgctgtg	gccttagtgg	gggcccaggt	ccctctccag	ccccagtctc	8820
	ctccttctgt	ccaatgagaa	agctgggatc	aggggtccct	gaggcccctg	tccactctgc	8880
15	atgcctcgat	ggtgaagctc	tgttggtatg	gcagagggga	ggctgctcag	gcatctgcat	8940
	ttcccctgcc	aatctagagg	atgaggaaag	ctctcaggaa	tagtaagcag	aatgtttgcc	9000
	ctggatgaat	aactgagctg	ccaat.taaca	aggggcaggg	agccttagac	agaaggtacc	9060
	aaatatgcct	gatgctccaa	cattttattt	gtaatatcca	agacaccctc	aaataaacat	9120
	atgattccaa	taaaaatgca	cagccacgat	ggcatctctt	agcctgacat	cgccacgatg	9180
20	tagaaattct	gcatcttcct	ctagttttga	attatcccca	cacaatcttt	ttcggcagct	9240
	tggatggtca	gtttcagcac	cttttacaga	tgatgaagct	gagcctcgag	ggatgtgtgt	9300
	cgtcaagggg	gctcagggct	tctcagggag	gggactcatg	gtttcttatt	ctgctacact	9360
	cttccaaacc	ttcactcacc	cctggtgatg	cccaccttcc	cctctctcca	ggcaaatggg	9420
	agagaccctt	tgaagtcaag	gacaccgagg	aagaggactt	ccacgtggac	caggtgacca	9480
25	ccgtgaaggt	gcctatgatg	aagcgtttag	gcatgtttaa	catccagcac	tgtaagaagc	9540
	tgtccagctg	ggtgctgctg	atgaaatacc	tgggcaatgc	caccgccatc	ttcttcctgc	9600
	ctgatgaggg	gaaactacag	cacctggtaa	atgaactcac	ccacgatatc	atcaccaagt	9660
			aggtgattcc			_	
20			actgaggccc				
30			actgctactc				
			cgtccatgct				
			gaacttccca				
			cagcagcacc			•	
25			ggcagcagtg	-			
35			tgtgtggcaa				
			aaggcttgcg				
			gatacttaaa				
		_	cctgctgatc		_	_	
4Ó			ggttcataat		•	•	
			tctgctccag				
			gagccagctg				
			cagagaggag				
-			tgggcactgc				
45			gggggccagg				
77	ggeereeagg		tatcagagga	ggtggcattt	caagettttg	cyaticatte	10/40



		• •				101/0	500/15/2
	gatgttaaca	ttctttgact	caatgtagaa	gagctaaaag	tagaacaaac	caaagccgag	10800
	ttcccatctt	agtgtgggtg	gaggacacag	gagtaagtgg	cagaaataat	cagaaaagaa	10860
	aacacttgca	ctgtggtggg	tcccagaaga	acaagaggaa	tgctgtgcca	tgccttgaat	10920
	ttcttttctg	cacgacaggt	ctgccagctt	acatttaccc	aaactgtcca	ttactggaac	10980
5	ctatgatctg	aagagcgtcc	tgggtcaact	gggcatcact	aaggtcttca	gcaatggggc	11040
	tgacctctcc	ggggtcacag	aggaggcacc	cctgaagctc	tccaaggtga	gatcaccctg	11100
	acgaccttgt	tgcaccatgg	tatctgtagg	gaagaatgtg	tgggggctgc	agcactgtcc	11160
	tgaggctgag	gaaggggccg	agggaaacaa	atgaagaccc	aggctgagct	cctgaagatg	11220
	cccgtgattc	actgacacyg	gacggtgggc	aaacagcaaa	gccaggcagg	ggctgctgtg	11280
10	cagctggcac	tttcggggcc	tcccttgagg	ttgtgtcact	gaccctgaat	ttcaactttg	11340
	cccaagacct	tctagacatt	gggccttgat	ttatccatac	tgacacagaa	aggtttgggc	11400
	taagttgttt	caaaggaatt	tctgactcct	tcgatctgtg	agatttggtg	tctgaattaa	11460
	tgaatgattt	cagctaaagt	gacacttatt	ttggaaaact	aaaggcgacc	aatgaacaac	11520
	ctgcagttcc	atgaatggct	gcattatctt	ggggtctggg	cactgtgaag	gtcactgcca	11580
15	gggtccgtgt	cctcaaggag	cttcaagccg	tgtactagaa	aggagagagc	cctggaggca	11640
	gacgtggagt	gacgatgctc	ttccctgttc	tgagttgtgg	gtgcacctga	gcagggggag	11700
	aggcgcttgt	caggaagatg	gacagagggg	agccagcccc	atcagccaaa	gccttgagga	11760
	ggagcaaggc	ctatgtgaca	gggagggaga	ggatgtgcag	ggccagggcc	gtccaggggg	11820
	agtgagcgct	tcctgggagg	tgtccacgtg	agccttgctc	gaggcctggg	atcagcctta	11880
20	caacgtgtct	ctgcttctct	cccctccagg	ccgtgcataa	ggctgtgctg	accatcgacg	11940
	agaaagggac	tgaagctgct	ggggccatgt	ttttagaggc	catacccatg	tctatccccc	12000
	ccgaggtcaa	gttcaacaaa	ccctttgtct	tcttaatgat	tgaacaaaat.	accaagtctc	12060
	ccctcttcat	gggaaaagtg	gtgaatccca	cccaaaaata	actgcctctc	gctcctcaac	12120
	ccctcccctc	catccctggc	cccctccctg	gatgacatta	aagaagggtt	gagctggtcc	12180
25	ctgcctgcat	gtgatctgta	aatccctggg	atgttttctc	tg		12222
	<210> 43	•					•
	-211- 1500						

<211> 1599

<212> DNA

30 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

35 <400> 43

40

45

caccagcate atetecteca atteatecag ctactetgee catgaagata atagtttea 60 ggeggattge cteagateae actateteea cttgeecage cetgtggaag attageggee 120 atgtateca atgtgatagg aactgtaace tetggaaaaa ggaaggttta tettttgtee 180 ttgetgetea ttggettetg ggaetgegtg acetgteaeg ggageectgt ggaeatetge 240 acageeaage egeggaeat teccatgaat eccatgtgea tttaeegete eeggagaag 300 aaggeaactg aggatgaggg eteagaacag aagateeegg aggeeaecaa eeggegtgte 360 tgggaactgt ecaaggeeaa tteeegett getaeeaett tetateagea eetggeagat 420 tecaagaatg acaatgataa eatttteetg teaeeectga gtateteeae ggettttget 480 atgaecaage tgggtgeetg taatgaeae eteegaaae teetgaega atetageae 600

	cgactctatc	gaaaagccaa	caaatcctcc	aagttagtat	cagccaatcg	cctttttgga	660
	gacaaatccc	ttaccttcaa	tgagacctac	caggacatca	gtgagttggt	atatggagcc	720
	aagctccagc	ccctggactt	caaggaaaat	gcagagcaat	ccagagcggc	catcaacaaa	780
	tgggtgtcca	ataagaccga	aggccgaatc	accgatgtca	ttccctcgga	agccatcaat	840
5	gagctcactg	ttctggtgct	ggttaacacc	atttacttca	agggcctgtg	gaagtcaaag	900
	ttcagccctg	agaacacaag	gaaggaactg	ttctacaagg	ctgatggaga	gtcgtgttca	960
	gcatctatga	tgtaccagga	aggcaagttc	cgttatcggc	gcgtggctga	aggcacccag	1020
	gtgcttgagt	tgcccttcaa	aggtgatgac	atcaccatgg	tcctcatctt	gcccaagcct	1080
	gagaagagcc	tggccaaggt	ggagaaggaa	ctcaccccag	aggtgctgca	ggagtggctg	1140
10	gatgaattgg	aggagatgat	gctggtggtt	cacatgcccc	gcttccgcat	tgaggacggc	1200
	ttcagtttga	aggagcagct	gcaagacatg	ggccttgtcg	atctgttcag	ccctgaaaag	1260
	tccaaactcc	caggtattgt	tgcagaaggc	cgagatgacc	tctatgtctc	agatgcattc	1320
	cataaggcat	ttcttgaggt	aaatgaagaa	ggcagtgaag	cagctgcaag	taccgctgtt	1380
	gtgattgctg	gccgttcgct	aaaccccaac	agggtgactt	tcaaggccaa	caggcccttc	1440
15	ctggttttta	taagagaagt	tcctctgaac	actattatct	tcatgggcag	agtagccaac	1500
	ccttgtgtta	agtaaaatgt	tcttattctt	tgcacctctt	cctatttttg	gtttgtgaac	1560
	agaagtaaaa	ataaatacaa	actacttcca	tctcacatt			1599
					•		
••	<210> 44						
20	<211> 23						
	<212> DNA						
	<213> T7 vi	rus					
25	<400> 44						
23	taatacgact	cactataggg	cga				23
	<210> 45		•				
	<211> 24						
	<212> DNA						
30	<213> T3 vi	rus				•	
	<400> 45						
	ttattaaccc	tcactaaagg	gaag	•			24
35	<210> 46						
	<211> 23						
	<212> DNA					•	
	<213> SP6 v	rirus					•
40	<400> 46	•					
	atttaggtga	cactatagaa	tac				22

<210> 47 <211> 555 <212> PRT <213> Mus musculus 5. <400> 47 Met Thr Lys Ser Ser Asn His Asn Cys Leu Leu Arg Pro Glu Asn Lys Pro Gly Leu Trp Gly Pro Gly Ala Gln Ala Ala Ser Leu Arg Pro Ser Pro Ala Thr Leu Val Val Ser Ser Pro Gly His Ala Glu His Pro Pro Ala Ala Pro Ala Gln Thr Pro Gly Pro Gln Val Ser Ala Ser Ala Arg 55 . Gly Pro Gly Pro Val Ala Gly Gly Ser Gly Arg Met Glu Arg Arg Met 75 · Lys Gly Gly Tyr Leu Asp Gln Arg Val Pro Tyr Thr Phe Cys Ser Lys Ser Pro Gly Asn Gly Ser Leu Gly Glu Ala Leu Met Val Pro Gln Gly Lys Leu Met Asp Pro Gly Ser Leu Pro Pro Ser Asp Ser Glu Asp Leu Phe Gln Asp Leu Ser His Phe Gln Glu Thr Trp Leu Ala Glu Ala Gln Val Pro Asp Ser Asp Glu Gln Phe Val Pro Asp Phe His Ser Glu Asn Leu Ala Phe His Ser Pro Thr Thr Arg Ile Lys Lys Glu Pro Gln Ser Pro Arg Thr Asp Pro Ala Leu Ser Cys Ser Arg Lys Pro Pro Leu Pro Tyr His His Gly Glu Gln Cys Leu Tyr Ser Arg Gln Ile Ala Ile Lys 

	Ser	Pro 210	Ala	Pro	Gly	Ala	Pro 215	Gly	Gln	Ser	Pro	Leu 220	Gln	Pro	Phe	Ser
5	Arg 225	Ala	Glu	Gln	Gln	Gln 230	Ser	Leu	Leu	Arg	Ala 235	Ser	Ser	Ser	Ser	Gln 240
	Ser	His	Pro	Gly	His 245	Gly	Tyr	Leu	Gly	Glu 250	His	Ser	Ser	Val	Phe 255	Gln
10	Gln	Pro	Val	Asp 260	Met	Cys	His	Ser	Phe 265	Thr	Ser	Pro	Gln	Gly 270	Gly	Gly
15	Arg	Glu	Pro 275	Leu	Pro	Ala	Pro	Tyr 280	Gln	His	Gln	Leu	Ser 285	Glu	Pro	Cys
	Pro	Pro 290	Tyr	Pro	Gln	Gln	Asn 295	Phe	Lys	Gln	Glu	Tyr 300	His	Asp	Pro	Leu
20	Tyr 305	Glu	Gln	Ala	Gly	Gln 310	Pro	Ala	Ser	Ser	Gln 315	Gly	Gly	Val	Ser	Gly 320
	His	Arg	Tyr	Pro	Gly 325	Ala	Gly	Val	Val	Ile 330	Lys	Gln	Glu	Arg	Thr 335	Asp
25			-	340		_			345	_				Tyr 350		
30			355	•		_		360		_	_		365	Met		
	-	370					375					380		Cys		
35	385			•		390					395			Gly		400
40					405					410			•	Leu	415	
40 .				420					425					Phe 430		
45	Ţrp	Thr	Gly 435	Arg	Gly	Met	Glu	Phe 440	Lys	Leu	Ile	Glu	Pro 445	Glu	Glu	Val

	Ala Arg Leu Trp Gly Ile Gln Lys Asn Arg Pro Ala Met Asn Tyr Asp 450 455 460	
5	Lys Leu Ser Arg Ser Leu Arg Tyr Tyr Tyr Glu Lys Gly Ile Met Gln 465 470 475 480	
	Lys Val Ala Gly Glu Arg Tyr Val Tyr Lys Phe Val Cys Glu Pro Glu 485 490 495	
10	Ala Leu Phe Ser Leu Ala Phe Pro Asp Asn Gln Arg Pro Ala Leu Lys 500 505 510	
15	Ala Glu Phe Asp Arg Pro Val Ser Glu Glu Asp Thr Val Pro Leu Ser 515 520 525	
13	His Leu Asp Glu Ser Pro Ala Tyr Leu Pro Glu Leu Thr Gly Pro Ala 530 535 540	
20	Pro Pro Phe Gly His Arg Gly Gly Tyr Ser Tyr 545 550 555	
25	<210> 48 <211> 8 <212> DNA <213> Artificial Sequence	
30	<220> <223> Description of Artificial Sequence: Synthetic <400> 48	
	cccggaag <210> 49	8
35	<211> 1843 <212> DNA <213> Homo sapiens	
. 40	<pre>&lt;400&gt; 49 ctgcaggggg ggggggggg gggggctgtc atggcggcag gacggcgaac ttgcagtatc tccacgaccc gcccctacag gtgccagtgc ctccagaatg tggcagctca caagcctcct gctgttcgtg gccacctggg gaatttccgg cacaccagct cctcttgact cagtgttctc</pre>	120 180
	cagcagegag egtgeecace aggtgetgeg gateegeaaa egtgeeaact eetteetgga ggageteegt cacageagee tggageggga gtgeatagag gagatetgtg aettegagga	

ggccaaggaa attttccaaa atgtggatga cacactggcc ttctggtcca agcacgtcga 360

45



	cggtgaccag	tgcttggtct	tgcccttgga	gcacccgtgc	gccagcctgt	gctgcgggcá	420
	cggcacgtgc	atcgacggca	tcggcagctt	cagctgcgac	tgccgcagcg	gctgggaggg	480
	ccgcttctgc	cagcgcgagg	tgagcttcct	caattgctcg	ctggacaacg	gcggctgcac	540
	gcattactgc	ctagaggagg	tgggctggcg	gcgctgtagc	tgtgcgcctg	gctacaagct	600
5	gggggacgac	ctcctgcagt	gtcaccccgc	agtgaagttc	ccttgtggga	ggccctggaa	660
	gcggatggag	aagaagcgca	gtcacctgaa	acgagacaca	gaagaccaag	aagaccaagt	720
	agatccgcgg	ctcattgatg	ggaagatgac	caggcgggga	gacagcccct	ggcaggtggt	780
	cctgctggac	tcaaagaaga	agctggcctg	cggggcagtg	ctcatccacc	cctcctgggt	840
	gctgacagcg	gcccactgca	tggatgagtc	caagaagctc	cttgtcaggc	ttggagagta	900
10	tgacctgcgg	cgctgggaga	agtgggagct	ggacctggac	atcaaggagg	tcttcgtcca	960
	ccccaactac	agcaagagca	ccaccgacaa	tgacatcgca	ctgctgcacc	tggcccagcc	1020
	cgccaccctc	tcgcagacca	tagtgcccat	ctgcctcccg	gacagcggcc	ttgcagagcg	1080
	cgagctcaat	caggccggcc	aggagaccct	cgtgacgggc	tggggctacc	acagcagccg	1140
	agagaa <u>g</u> gag	gccaagagaa	accgcacctt	cgtcctcaac	ttcatcaaga	ttcccgtggt	1200
15	cccgcacaat	gagtgcagcg	aggtcatgag	caacatggtg	tctgagaaca	tgctgtgtgc	1260
	gggcatcctc	ggggaccggc	aggatgcctg	cgagggcgac	agtggggggc	ccatggtcgc	1320
	ctccttccac	ggcacctggt	tcctggtggg	cctggtgagc	tggggtgagg	gctgtgggct	1380
	ccttcacaac	tacggcgttt	acaccaaagt	cagccgctac	ctcgactgga	tccatgggca	1440
	catcagagac	aaggaagccc	cccagaagag	ctgggcacct	tagcgaccct	ccctgcaggg	1500
20	ctgggctttt	gcatggcaat	ggatgggaca	ttaaagggac	atgtaacaag	cacaccggcc	1560
	tgctgttctg	tccttccatc	cctcttttgg	gctcttctgg	agggaagtaa	catttactga	1620
						caactctgtg	
	gggtggggag	gagcagatcc	aagttttgcg	gggtctaaag	ctgtgtgtgt	tgagggggat	1740
	actctgttta	tgaaaaagaa	taaaaaacac	aaccacgaaa	aaaaaaaaa	aaaaaaaaa	1800
25	aaaaaaaaa	aaaaaaaccc	cccccgccc	ccccccctg	cag		1843
		•			• .		
	<210> 50	•					
	<211> 11729	5					
20	<212> DNA						
30	<213> Homo	sapiens					
	<400> 50						
						agaggttaga	
25						gatttgtgag	
35						cccaaaaagg	
						ggacagcgtc	
						attttagatt	
						tatctctggg	
40						ccactatggg	
40 <sup>.</sup>		•				gctggatggc	
						aggaccagga	
						ccaagcaaat	
						ggcgaacttg	
4.5	cagtatctcc	acgacccgcc	cctgtgagtc	cccctccagg	caggtctatg	aggggtgtgg	720

agggagggct gccccggga gaagagagct aggtggtgat gagggctgaa tcctccagcc 780

45

	agggtgctca	äcaagcctga	gcttggggta	aaaggacaca	aggccctcca	caggccaggc	840
	ctggcagcca	cagtctcagg	tccctttgcc	atgcgcctcc	ctctttccag	gccaagggtc	900
	cccaggccca	gggccattcc	aacagacagt	ttggagccca	ggaccctcca	ttctccccac	960
	cccacttcca	cctttggggg	tgtcggattt	gaacaaatct	cagaagcggc	ctcagaggga	1020
5 ·	gtcggcaaga	atggagagca	gggtccggta	gggtgtgcag	aggccacgtg	gcctatccac	1080
	tggggagggt	tccttgatct	ctggccacca	gggctatctc	tgtggccttt	tggagcaacc	1140
	tggtggtttg	gggcaggggt	tgaatttcca	ggcctaaaac	cacacaggcc	tggccttgag	1200
	tcctggctct	gcgagtaatg	catggatgta	aacatggaga	cccaggacct	tgcctcagtc	1260
	ttccgagtct	ggtgcctgca	gtgtactgat	ggtgtgagac	cctactcctg	gaggatgggg	1320
10	gacagaatct	gatcgatccc	ctgggttggt	gacttccctg	tgcaatcaac	ggagaccagc	1380
	aagggttgga	tttttaataa	accacttaac	tcctccgagt	ctcagtttcc	ccctctatga	1440
	aatggggttg	acagcattaa	taactacctc	ttgggtggtt	gtgagcctta	actgaagtca	1500
	taatatctca	tgtttactga	gcatgagcta	tgtgcaaagc	ctgttttgag	agctttatgt	1560
	ggactaactc	ctttaattct	cacaacaccc	tttaaggcac	agatacacca	cgttattcca	1620
15	tccattttac	aaatgaggaa	actgaggcat	ggagcagtta	agcatcttgc	ccaacattgc	1680
	cctccagtaa	gtgctggagc	tggaatttgc	accgtgcagt	ctggcttcat	ggcctgccct	1740
	gtgaatcctg	taaaaattgt	ttgaaagaca	ccatgagtgt	ccaatcaacg	ttagctaata	1800
	ttctcagccc	agtcatcaga	ccggcagagg	cagccacccc	actgtcccca	gggaggacac	1860
••			cactgcattc			_	
20			gcagccgagg				
			gcctgacggc				
			agagcagcag				
			tgcctccaga				
25			cggcacacca				
25			ctacaaggcc				
			tgggtgaggt				
			agctgccaga				
•			gcgatgacct				
30			tccctctcca				
30			actatccagc				
			acccctttcc				
			ggcacatggg				
			gtttatctgc ccctcctcta			_	
35			ccctcctaac				
			gacccttcct				
			atggggaata				
			gcatgtggtg				
			ccagggcttg			_	
40			atccccatgt				
			tcacacagag				
			tcctccaact				
			atgacctaga			-	
			agctctgctt				
45			agctcccctc	•			
					_	~	



	agcgagcgtg	cccaccaggt	gctgcggatc	cgcaaacgtg	ccaactcctt	cctggaggag	3540
	ctccgtcaca	gcagcctgga	gcgggagtgc	atagaggaga	tctgtgactt	cgaggaggcc	3600
	aaggaaattt	tccaaaatgt	ggatgacaca	gtaaggccac	catgggtcca	gaggatgagg	3660
	ctcaggggcg	agctggtaac	cagcaggggc	ctcgaggagc	aggtggggac	tcaatgctga	3720
5	ggccctctta	ggagttgtgg	gggtggctga	gtggagcgat	taggatgctg	gccctatgat	3780
	gtcggccagg	cacatgtgac	tgcaagaaac	agaattcagg	aagaagctcc	aggaaagagt	3840
	gtggggtgac	cctaggtggg	gactcccaca	gccacagtgt	aggtggttca	gtccaccctc	3900
	cagccactgc	tgagcaccac	tgcctccccg	tcccacctca	caaagagggg	acctaaagac	3960
	caccctgctt	ccacccatgc	ctctgctgat	cagggtgtgt	gtgtgaccga	aactcacttc	4020
10	tgtccacata	aaatcgctca	ctctgtgcct	cacatcaaag	ggagaaaatc	tgattgttca	4080
	gggggtcgga	agacagggtc	tgtgtcctat	ttgtctaagg	gtcagagtcc	tttggagccc	4140
	ccagagtcct	gtggacgtgg	ccctaggtag	tagggtgagc	ttggtaacgg	ggctggcttc	4200
	ctgagacaag	gctcagaccc	gctctgtccc	tggggatcgc	ttcagccacc	aggacctgaa	4260
	aattgtgcac	gcctgggccc	ccttccaagg	catccaggga	tgctttccag	tggaggcttt	4320
15	cagggcagga	gaccctctgg	cctgcaccct	ctcttgccct	cagcctccac	ctccttgact	4380
	ggacccccat	ctggacctcc	atccccacca	cctctttccc	cagtggcctc	cctggcagac	4440
	accacagtga	ctttctgcag	gcacatatct	gatcacatca	agtccccacc	gtgctcccac	4500
	ctcacccatg	gtctctcagc	cccagcagcc	ttggctggcc	tctctgatgg	agcaggcatc	4560
	aggcacaggc	cgtgggtctc	aacgtgggct	gggtggtcct	ggaccagcag	cagccgccgc	4620
20	agcagcaacc	ctggtacctg	gttaggaacg	cagaccctct	gcccccatcc	tcccaactct	4680
						ccgcaggcag	
			•			gtcccttgca	
						catggggaga	
0.5						ctggagggcg	
25						cgatccctgt	
						acccgggcgc	
						agctggcctt	
						ggcgcccgcg	
20						cggtgaccag	
30						cggcacgtgc	
						ccgcttctgc	
						ggggccgggt	
						ggtgagcttc	
35						ggtgggctgg	
33						gtgtcacccc	
						gtgggcaggc	
						gaggaacaga	
					_	gcgcaatcag	
40						gtccccgctt	
70						tggtggctcc	
						tgtagcctgg	
						ccccttgctt	
						tttatgcatt	
45						tcttacactt	
73	ccagcatgct	gccccttggc	acgggtcctt	ttttcattca	ccccataaa	aggtggaccc	<b>6180</b>

WO 00/75279



	ttttaatgtg	gaaattccta	tcttctgcct	ctagggcatt	tatcacttat	ttcttctaca	6240
	atctcccctt	tacttcctct	attttctctt	tctggacctc	ccattattca	gacctctttc	6300
	ctctagtttt	attgtctctt	ctatttccca	tctctttgac	tttgtgtttt	ctttcaggga	6360
	actttcttt	ttttctttt	ttttgagatg	gagtttcact	cttgttgtcc	caggctggag	6420
5	tgcaatgacg	tgatctcagc	tcaccacaac	ctccgcctcc	tggattcaag	cgattctcct	6480
	gccgcagcct	cccgagtagc	tgggattaca	ggcatgcgcc	accacgccca	gctaattttg	6540
	tgtttttagt	agagaagggg	tttctccgtg	ttggtcaagc	tggtcttgaa	ctcctgacct	6600
	caggtgatcc	acctgccttg	gcctcctaaa	gtgctgggat	tacaggcgtg	agccaccgcg	6660
	cccagcctct	ttcagggaac	tttctacaac	tttataattc	aattcttctg	cagaaaaaaa	6720
10	tttttggcca	ggctcagtag	ctcagaccaa	taattccagc	actttgagag	gctgaggtgg	6780
	gaggattgct	tgagcttggg	agtttgagac	tagcctgggc	aacacagtga	gaccctgtct	6840
	ctatttttaa	aaaaagtaaa	aaaagatcta	aaaatttaac	tttttattt	gaaataatta	6900
	gatatttcca	ggaagctgca	aagaaatgcc	tggtgggcct	gttggctgtg	ggtttcctgc	6960
	aaggccgtgg	gaaggccctg	tcattggcag	aaccccagat	cgtgagggct	ttccttttag	7020
15	gctgctttct	aagaggactc	ctccaagctc	ttggaggatg	gaagacgctc	acccatggtg	7080
	ttcggcccct	cagagcaggg	tggggcaggg	gagctggtgc	ctgtgcaggc	tgtggacatt	7140
	tgcatgactc	cctgtggtca	gctaagagca	ccactccttc	ctgaagcggg	gcctgaagtc	7200
	cctagtcaga	gcctctggtt	caccttctgc	aggcagggag	aggggagtca	agtcagtgag	7260
	gagggctttc	gcagtttctc	ttacaaactc	tcaacatgcc	ctcccacctg	cactgccttc	7320
20	ctggaagccc	cacagcctcc	tatggttdcg	tggtccagtc	cttcagcttc	tgggcgcccc	7380
	catcacgggc	tgagatttt	gctttccagt	ctgccaagtc	agttactgtg	tccatccatc	7440
	tgctgtcagc	ttctggaatt	gttgctgttg	tgccctttcc	attcttttgt	tatgatgcag	7500
	ctcccctgct	gacgacgtcc	cattgctctt	ttaagtctag	atatctggac	tgggcattca	7560
	aggcccattt	tgagcagagt	cgggctgacc	tttcagccct	cagttctcca	tggagtatgc	7620
25	getetettet	tggcagggag	gcctcacaaa	catgccatgc.	ctattgtagc	agctctccaa	7680
	gaatgctcac	ctccttctcc	ctgtaattcc	tttcctctgt	gaggagctca	gcagcatccc	7740
	attatgagac	cttactaatc	ccagggatca	ccccaacag	ccctgggġta	caatgagctt	7800
	ttaagaagtt	taaccaccta	tgtaaggaga	cacaggcagt	gggcgatgct	gcctggcctg	7860
	actcttgcca	ttgggtggta	ctgtttgttg	actgactgac	tgactgactg	gagggggttt	7920
30	gtaatttgta	tctcagggat	tacccccaac	agccctgggg	tacaatgagc	cttcaagaag	7980
	tttaacaacc	tatgtaagga	cacacagcca	gtgggtgatg	ctgcctggtc	tgactcttgc	8040
	cattcagtgg	çactgtttgt	tgactgactg	actgactgac	tggctgactg	gaggggttc	8100
	atagctaata	ttaatggagt	ggtctaagta	tcattggttc	cttgaaccct	gcactgtggc	8160
	aaagtggccc	acaggctgga	ggaggaccaa	gacaggaggg	cagtctcggg	aggagtgcct	8220
35	ggcaggcccc	tçaccacctc	tgcctacctc	agtgaagttc	ccttgtggga	ggccctggaa	8280
	gcggatggag	aagaagcgca	gtcacctgaa	acgagacaca	gaagaccaag	aagaccaagt	8340
					gacagcccct		
	aggcgaggca	gcaccggctc	gtcacgtgct	gggtccggga	tcactgagtc	catcctggca	8460
					attgcgtttg		
40	aaggtggggg	atgetteagg	gaaagatgga	cgcaacctga	ggggagagga	gcagccaggg	8580
					ggagggaggg	-	
					taggcagaag	_	
					ccagcctcca		
45					ataggaacca		
45	gtatttgccc	tggggactca	gactctgcaa	gggtcaggac	cccaaagacc	cggcagccca	8880

	gtgggaccac	agccaggacg	gcccttcaag	ataggggctg	agggaggcca	aggggaacat	8940
	ccaggcagcc	tgggggccac	aaagtcttcc	tggaagacac	aaggcctgcc	aagcctctaa	9000
	ggatgagagg	agctcgctgg	gcgatgttgg	tgtggctgag	ggtgactgaa	acagtatgaa	9060
	cagtgcagga	acagcatggg	caaaggcagg	aagacaccct	gggacaggct	gacactgtaa	9120
5	aatgggcaaa	aatagaaaac	gccagaaagg	cctaagccta	tgcccatatg	accagggaac	9180
•	ccaggaaagt	gcatatgaaa	cccaggtgcc	ctggactgga	ggctgtcagg	aggcagccct	9240
	gtgatgtcat	catcccaccc	cattccaggt	ggtcctgctg	gactcaaaga	agaagctggc	9300
	ctgcggggca	gtgctcatcc	acccctcctg	ggtgctgaca	gcggcccact	gcatggatga	9360
	gtccaagaag	ctccttgtca	ggcttggtat	gggctggagc	caggcagaag	ggggctgcca	9420
10	gaggcctggg	tagggggacc	aggcaggctg	ttcaggtttg	ggggaccccg	ctccccaggt	9480
	gcttaagcaa	gaggcttctt	gagctccaca	gaaggtgttt	gggggaaga	ggcctatgtg	9540
	ccccaccct	gcccacccat	gtacacccag	tattttgcag	tagggggttc	tctggtgccc	9600
	tcttcgaatc	tgggcacagg	tacctgcaca	cacatgtttg	tgaggggcta	cacagacctt	9660
	cacctctcca	ctcccactca	tgaggagcag	gctgtgtggg	cctcagcacc	cttgggtgca	9720
15	gagaccagca	aggcctggcc	tcagggctgt	gcctcccaca	gactgacagg	gatggagctg	9780
	tacagaggga	gccctagcat	ctgccaaagc	cacaagctgc	ttccctagca	ggctgggggc	9840
	tcctatgcat	tggccccgat	ctatggcaat	ttctggaggg	ggggtctggc	tcaactcttt	9900
	atgccaaaaa	gaaggcaaag	catattgaga	aaggccaaat	tcacatttcc	tacagcataa	9960
	tctatgccag	tggccccgtg	gggcttggct	tagaattccc	aggtgctctt	cccagggaac	10020
20	catcagtctg	gactgagagg	accttctctc	tcaggtggga	cccggccctg	tcctccctgg	10080
	cagtgccgtg	ttctgggggt	cctcctct	gggtctcact	gcccctgggg	tctctccagc	10140
	tacctttgct	ccatgttcct	ttgtggctct	ggtctgtgtc	tggggtttcc	aggggtctcg	10200
	ggcttccctg	ctgcccattc	cttctctggt	ctcacggctc	cgtgactcct	gaaaaccaac	10260
	cagcatccta	cccctttgga	ttgacacctg	ttggccactc	cttctggcag	gaaaagtcac	10320
25		gttccacggc					
	tgcacagtct	ccgggtgaac	cttcttcagg	ccctctccca	ggcctgcagg	ggcacagcag	10440
	tgggtgggcc	tcaggaaagt	gccactgggg	agaggctccc	cgcagcccac	tctgactgtg	10500
		tgcaggagag		•			
20		ggtcttcgtc		·			
30		cctggcccag					
	•	ccttgcagag					
		ccacagcagc					
		gattcccgtg					
25		catgctgtgt		14			
35		gcccatggtc					
		gggctgtggg				_	
		gatccatggg					
		ctccctgcag					
40		agcacaccgg					
40		aacatttact					
		agcaactctg					
		gttgagggg				=	
	_	gccttttcca					
45		accagettte					-
43	cacagaggag	gaaactgagg	ggtctgaaag	gtttacatgg	tggagccagg	attcaaatct	11580

aggtctgact ccaaaaccca ggtgcttttt tctgttctcc actgtcctgg aggacagctg 11640 tttcgacggt gctcagtgtg gaggccacta ttagctctgt agggaagcag ccagagaccc 11700 agaaagtgtt ggttcagccc agaat 11725 5 <210> 51 <211> 32 <212> DNA <213> Artificial Sequence 10 <220> <223> Description of Artificial Sequence: Synthetic <400> 51 ttattttata tatataatat atatataaaa ta 32 15 <210> 52 <211> 9 <212> DNA <213> Artificial Sequence 20 <220> <223> Description of Artificial Sequence: Synthetic <400> 52 25 tataatata 9 <210> 53 <211> 16 <212> DNA 30 <213> Artificial Sequence <223> Description of Artificial Sequence: Synthetic 35 <400> 53 caatataaat atatag 16 <210> 54 <211> 44 40 <212> DNA <213> Artificial Sequence

<223> Description of Artificial Sequence: Synthetic

<220>

45

	WO 00/73279	PC1/US00/15/2
	<400> 54	
	tgtgtgtgta tgcgtgtgtg tagacacaca cgcatacaca cata	44
	<210> 55	
5	<211> 41	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
10	<223> Description of Artificial Sequence: Synthetic	
	<400> 55	
	ttattttata tatataatat atataaaaa tatataata	41
15	<210> 56	
	<211> 25	
	<212> DNA	
	<213> Artificial Sequence	
20	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 55	
25	tataatatac aatataaata tatag	25
23	<210> 57	
	<211> 60	•
	<212> DNA	
	<213> Artificial Sequence	
30		
	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 57	
35	caatataaat atatagtgtg tgtgtatgcg tgtgtgtaga cacacacgca tac	acacata 60
	<210> 58	
	<211> 101	
	<212> DNA	
40	<213> Artificial Sequence	
	· <220>	

<223> Description of Artificial Sequence: Synthetic



	<400> 58						
		tatataatat	atatataaaa	tatataatat	acaatataaa	tatatagtgt	60
				atacacacat		cacacagogo	101
	5-5-5			acacacacac	<u> </u>		101
5	<210> 59						
	<211> 723						
	<212> DNA						
	<213> Homo	sapiens					•
		•				•	
10	<400> 59						
•	tttaggaaaa	ttgttaagac	tgggttgtat	gcacactggt	gtttattata	ttatgtagtt	60
	ttttctgtat	ttttacaaca	tttcagaatt	aaaagcaaca	gctagaaaaa	gagggaaatg	120
	gccgggtgca	gtcgtcacgc	ctgtaatccc	agcactttgg	gaggccaagg	cgggcggatc	180
	acgaggtcga	gagatcgaga	ccatcctggc	caacatggtg	aaaccccatc	tctactaaaa	240
15	atacaaaaat	taactgggca	tggtggcatg	cgcctgtagt	cccaggagaa	ttgcttgaac	300
	ctgggaggcg	gaggttgcag	tgagccaaga	tctcaccact	gctctccagc	ctggtgacag	360
	ggcaagactc	cgtcaaaaaa	aaaaaagaga	gggagagcca	gagtatgaaa	aaggaagtca	420
	gagcccttta	atgagtcagc	tttgtaggtc	tccaggtagg	aggctagtgc	ttcagtgtct	480
	aggacatagt	aggtgttcag	taaattaaat	tcaggacaaa	aagaacatgc	cccaaggacc	540
20	atctgatatc	cacttaaagt	gatggactac	ctcgtttccc	ttgtttatga	atgggttcat.	600
•	gcctaagact	gtgtgcactt	taatacaagg	gcagtcgttc	agaactagtc	aggtcctgaa	660
	aaggatttac	caaatgttga	gtgtgccctc	tagtgttcac	acttcccagc	tttcttccta	720
	taa						723
					,		
25	<210> 60					•	
	<211> 447						
	<212> DNA						
	<213> Homo	sapiens					
20			•			•	
30	<400> 60			•			
						tcaatgcgtc	
•						taggacaáag	
·						gaaagaatga	
25						tggaggcctt	
35						tagaaaaatc	
						aggtacaggg	
				tcaagccagt	gtgtttccta	atgtatctgc	420
	tgtatcccca	tgaagcaaat	tttgcca				447
40							
40	<210> 61						
	<211> 103						

<212> DNA

<213> Homo sapiens

	<400> 61	
	aggtggatca aggcacttgc ttacaactgg aactgaaatc ctccaagtcg atctagacat	60
	tgagatggag aaaatattca ttgtcgactg taattatgca acg	103
5	<210> 62	
	<211> 7	
	<212> DNA	
	<213> Artificial Sequence	
10	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 62	
	gtggaag	7
15		
	<210> 63	
	<211> 7	•
	<212> DNA	
	<213> Artificial Sequence	
20		
	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 63	
25	gtggatg	7
	·	
	<210> 64 .	
	<211> 7	
	<212> DNA	
30	<213> Artificial Sequence	
	<220>	
	<223> Description of Artificial Sequence: Synthetic	
35	<400> 64	
	ctggaag	7
	<210> 65	
	<211> 7	
40	<212> DNA	
	<213> Artificial Sequence	
	<220>	
4.5	<223> Description of Artificial Sequence: Synthetic	
45		

	WO 00/75279		PCT/US00/15728
	<400> 65		•
	ctggatg		7
	<210> 66		
5	<211> 32		
	<212> DNA		
	<213> Artificial Sequence		
	<220>		•
10	<pre>&lt;223&gt; Description of Artificial Sequence:</pre>	Synthetic	
	<400> 66		
	acccattcag tcgaggaagg atagggtggt at		32
15	<210> 67	•	
	<211> 32		
	<212> DNA		•
	<213> Artificial Sequence		
20	<220>		•
	<223> Description of Artificial Sequence:	Synthetic	
	<400> 67		
	agccattgag tcgaggaagg atagggtggt at		32
25			
	<210> 68		
	<211> 32		
	<212> DNA		
	<213> Artificial Sequence	•	
30	·		
	<220>		
	<223> Description of Artificial Sequence:	Synthetic	
	<400> 68	•	
35	agccattcag acgaggaagg atagggtggt at		32
	<210> 69		
	<211> 32		
	<212> DNA		
40	<213> Artificial Sequence		•
	<220>		
•	<223> Description of Artificial Sequence:	Synthetic	

	WO 00/75279	(	PCT/US00/15728
	<400> 69		•
	agccattcag tçgaggaagg atagggtggt tt		. 32
	<210> 70		
5	<211> 32		
	<212> DNA		
	<213> Artificial Sequence		
•	<220>		
10	<223> Description of Artificial Sequence:	Synthetic	·
	<400> 70		
	agccattcag tcgaggatcc caagggtggt at		32
15	<210> 71		
	<211> 32		
	<212> DNA		
	<213> Artificial Sequence		
20	<220>		
	<223> Description of Artificial Sequence:	Synthetic	
	<400> 71		
	agccattcag tcgaggaagg atagggccta at		32 .
25			
	<210> 72		
	<211> 33		
	<212> DNA		
20	<213> Artificial Sequence	•	
30	222		
•	<220> calculation of Artificial Sequence:	Cimthotic	
	(223) Description of Artificial Sequence:	synchecic	
	<400> 72		
35	agaccattca gtcgaggaag gatagggtgg tat		33
	<210> 73		
	<211> 33		
	<212> DNA		
. 40	<213> Artificial Sequence		
	<220>		
ı	<223> Description of Artificial Sequence:	Synthetic	

	WO 00/75279	PCT/US00/15728
	<400> 73	
	agccattcag tcgaggaagg atagcggtgg tat	33
	<210> 74	
5	<211> 25	
	<212> DNA	
	<213> Artificial Sequence	
	<220>	
10	<223> Description of Artificial Sequence: Synthetic	
	<400> 74	
	agccattcag tcgaggaagg ataat	25
15	<210> 75	
	<211> 28	
	<212> DNA	
	<213> Artificial Sequence	
20	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 75	
	agccattcga ggaaggatag ggtggtat	28
25	•	
•	<210> 76	
٠	<211> 1273	
	<212> DNA	•
30	<213> Artificial Sequence	
	<220>	
	<223> Description of Artificial Sequence: Synthetic	
	<400> 76	
35	ggcctctgac taactaatca ctttcccatc ttttgttaga tttgaatata tac	attctat 60
	gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaa	ttgatta 120
	gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tct	aagggcc 180
	cagcccttga caaaattgtg aagttaaatt ctccactctg tccatcagat act	atggttc 240
	tccactatgg caactaactc actcaatttt ccctccttag cagcattcca tct	tcccgat 300
40	cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gta	caggatc 360
	tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca cag	
	tgagaggcta aaactcatca aaaacactac teetttteet etacectatt eet	
	ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cct	
4.5	ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tac	<u>=</u>
45	cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tqc	ttttcaq 660

```
aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaaqtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatatag 780
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
 5
      agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
      aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
10
      aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttaatccttc 1260
      tatcttgaat ctt
                                                                         1273
      <210> 77
      <211> 1273
15
      <212> DNA
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
20
      <400> 77
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
25
      cagcccttga caaaattgtg aagttaaatt ctccactctg tccatcagat actatggttc 240
      tccactatgg caactaactc actcaaattt ccctccttag cagcattcca tcttcccgat 300
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggcta aaactcatca aaaacactac tccttttcct ctaccctatt cctcaatctt 480
30
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
      ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600
      cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatatag 780
35
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
      agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
40
      aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
      aagttgtoot titotggttt ogtgttoaco atggaacatt ttgattatag ttaatootto 1260
```

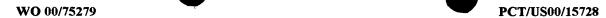
1273

tatcttgaat ctt

```
<210> 78
       <211> 1273
      <212> DNA
       <213> Artificial Sequence
. 5
       <220>
      <223> Description of Artificial Sequence: Synthetic
       <400> 78
10
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
      cagcccttga caaaattgtg aagttaaatt ctccactctg tccatcagat actatggttc 240
      tecaetatgg caactaacte acteaatttt cecteettag cageatteca tettecegat 300
15
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggcta aaactcatca aaaacactac tccttttcct ctaccctatt cctcaatctt 480
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
      ttttaccctc catggtcgtt aaaggagaga tggggacgat cattctgtta tacttctgta 600
20
      cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatatag 780
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
       agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
25
      agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
       aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
30
      aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttaatccttc 1260
      tatcttgaat ctt
                                                                         1273
      <210> 79
       <211> 1273
35
      <212> DNA
      <213> Artificial Sequence
       <223> Description of Artificial Sequence: Synthetic
40
       <400> 79
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
45
      cagecettga caaaattgtg aagttaaatt etecaetetg tecateagat actatggtte 240
```



```
tccactatgg caactaactc actcaatttt ccctccttag cagcattcca tcttcccqat 300
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggeta aaacteatea aaaacaetae teetttteet etaceetatt eeteaatett 480
 5
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
      ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600
      cacagitata catgictate aaacccagae tigeticeat aqiqqaqaet tqettiteaq 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatatag 780
10
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
      agattggcat atcattgtaa ctaaaaaaqc tqacattqac ccaqacatat tqtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
15
      aagaagttga agttgcctag accagaggac ataagtatgt actctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
      aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttaatccttc 1260
      tatcttgaat ctt
20
      <210> 80
      <211> 1275
      <212> DNA
      <213> Artificial Sequence
25
      <220>
      <223> Description of Artificial Sequence: Synthetic
      <400> 80
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
30
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
      cagecettga caaaattgtg aagttaaatt etecaetetg tecateagat actatggtte 240
      tecaetatgg caactaacte acteaatttt cecteettag cageatteca tettecegat 300
      ttcttctttg cttctccaac caaaacatca atgtttatta gttctgtata cagtacagga 360
35
      tctttggtct actctatcac aaggccagta ccacactcat gaagaaagaa cacaggagta 420
      gctgagaggc taaaactcat caaaaacact actccttttc ctctacccta ttcctcaatc 480
      ttttaccttt tccaaatccc aatccccaaa tcagtttttc tctttcttac tccctctctc 540
      ccttttaccc tccatggtcg ttaaaggaga gatggggagc atcattctgt tatacttctg 600
      tacacagtta tacatgtcta tcaaacccag acttgcttcc atagtggaga cttgctttc 660
40
      agaacatagg gatgaagtaa ggtgcctgaa aagtttgggg gaaaagtttc tttcagagag 720
      ttaagttatt ttatatatat aatatatat taaaatatat aatatacaat ataaatatat 780
      agtgtgtgtg tgtatgcgtg tgtgtagaca cacacgcata cacacatata atggaagcaa 840
      taagccattc taagagcttg tatggttatg gaggtctgac taggcatgat ttcacgaagg 900
      caagattggc atatcattgt aactaaaaaa gctgacattg acccagacat attgtactct 960
45
      ttctaaaaat aataataata atgctaacag aaagaagaga accgttcgtt tgcaatctac 1020
```



```
agctagtaga gactttgagg aagaattcaa cagtgtgtct tcagcagtgt tcagagccaa 1080
      gcaagaagtt gaagttgcct agaccagagg acataagtat catgtctcct ttaactagca 1140
      taccccgaag tggagaaggg tgcagcaggc tcaaaggcat aaqtcattcc aatcagccaa 1200
      ctaagttgtc cttttctggt ttcgtgttca ccatggaaca ttttgattat agttaatcct 1260
 5
      tctatcttga atctt
                                                                         1275
      <210> 81
      <211> 1276
      <212> DNA
10
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
15
      <400> 81
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttttctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
      cagcccttga caaaattgtg aagttaaatt ctccactctg tccatcagat actatggttc 240
20
      tccactatgg caactaactc actcaatttt ccctccttag cagcattcca tcttcccgat 300
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggcta aaactcatca aaaacactac teetttteet etaceetatt eeteaatett 480
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
25
      ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600
      cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatataq 780
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
30
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
      agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
      aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
35
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
      gctaagttgt ccttttctgg tttcgtgttc accatggaac attttgatta tagttaatcc 1260
      ttctatcttg aatctt
                                                                         1276
      <210> 82
40
      <211> 1272
      <212> DNA
```

<213> Artificial Sequence

<220>

45

<223> Description of Artificial Sequence: Synthetic

```
<400> 82
5
      ggcctctcac taactaatca ctttcccatc ttttgttaga tttgaatata tacattctat 60
      gatcattgct ttitctcttt acaggggaga atttcatatt ttacctgagc aaattgatta 120
      gaaaatggaa ccactagagg aatataatgt gttaggaaat tacagtcatt tctaagggcc 180
      cagcccttga caaaattgtg aagttaaatt ctccactctg tccatcagat actatggttc 240
      tocactatgg caactaacto actoaatttt coctoottag cagoattoca tottocogat 300
10
      cttctttgct tctccaacca aaacatcaat gtttattagt tctgtataca gtacaggatc 360
      tttggtctac tctatcacaa ggccagtacc acactcatga agaaagaaca caggagtagc 420
      tgagaggeta aaacteatea aaaacactae teetttteet etaceetatt eeteaatett 480
      ttaccttttc caaatcccaa tccccaaatc agtttttctc tttcttactc cctctctccc 540
      ttttaccctc catggtcgtt aaaggagaga tggggagcat cattctgtta tacttctgta 600
15
      cacagttata catgtctatc aaacccagac ttgcttccat agtggagact tgcttttcag 660
      aacataggga tgaagtaagg tgcctgaaaa gtttggggga aaagtttctt tcagagagtt 720
      aagttatttt atatatataa tatatatata aaatatataa tatacaatat aaatatatag 780
      tgtgtgtgtg tatgcgtgtg tgtagacaca cacgcataca cacatataat ggaagcaata 840
      agccattcta agagcttgta tggttatgga ggtctgacta ggcatgattt cacgaaggca 900
20
      agattggcat atcattgtaa ctaaaaaagc tgacattgac ccagacatat tgtactcttt 960
      ctaaaaataa taataataat gctaacagaa agaagagaac cgttcgtttg caatctacag 1020
      ctagtagaga ctttgaggaa gaattcaaca gtgtgtcttc agcagtgttc agagccaagc 1080
      aagaagttga agttgcctag accagaggac ataagtatca tgtctccttt aactagcata 1140
      ccccgaagtg gagaagggtg cagcaggctc aaaggcataa gtcattccaa tcagccaact 1200
25
      aagttgtcct tttctggttt cgtgttcacc atggaacatt ttgattatag ttatccttct 1260
      atcttgaatc tt
                                                                         1272
      <210> 83
      <211> 1272
30
      <212> DNA
      <213> Artificial Sequence
      <220>
      <223> Description of Artificial Sequence: Synthetic
35
      <400> 83
      gcctctcact aactaatcac tttcccatct tttgttagat ttgaatatat acattctatg 60
      atcattgctt tttctcttta caggggagaa tttcatattt tacctgagca aattgattag 120
      aaaatggaac cactagagga atataatgtg ttaggaaatt acagtcattt ctaagggccc 180
40
      agcccttgac aaaattgtga agttaaattc tccactctgt ccatcagata ctatggttct 240
```

ccactatggc aactaactca ctcaattttc cctccttagc agcattccat cttcccgatc 300 ttctttgctt ctccaaccaa aacatcaatg tttattagtt ctgtatacag tacaggatct 360 ttggtctact ctatcacaag gccagtacca cactcatgaa gaaagaacac aggagtagct 420 gagaggctaa aactcatcaa aaacactact ccttttcctc taccctattc ctcaatcttt 480

taccttttcc aaatcccaat ccccaaatca gtttttctct ttcttactcc ctctctccct 540

	tttaccctcc	atggtcgtta	aaggagagat	ggggagcatc	attctgttat	acttctgtac	600.
	acagttatac	atgtctatca	aacccagact	tgcttccata	gtggagactt	gcttttcaga	660
	acatagggat	gaagtaaggt	gcctgaaaag	tttgggggaa	aagtttcttt	cagagagtta	720
	agttatttta	tatatataat	atatatata	aatatataat	atacaatata	aatatatagt	780
5	gtgtgtgtgt	atgcgtgtgt	gtagacacac	acgcatacac	acatataatg	gaagcaataa	840
	gccattctaa	gagcttgtat	ggttatggag	gtctgactag	gcatgatttc	acgaaggcaa	900
	gattggcata	tcattgtaac	taaaaaagct	gacattgacc	cagacatatt	gtactctttc	960
	taaaaataat	aataataatg	ctaacagaaa	gaagagaacc	gttcgtttgc	aatctacagc	1020
	tagtagagac	tttgaggaag	aattcaacag	tgtgtcttca	gcagtgttca	gagccaagca	1080
10	agaagttgaa	gttgcctaga	ccagaggaca	taagtatcat	gtctccttta	actagcatac	1140
	cccgaagt <u>g</u> g	agaagggtgc	agcaggctca	aaggcataag	tcattccaat	cagccaacta	1200
	agttgtcctt	ttctggtttc	gtgttcacca	tggaacattt	tgattatagt	taatccttct	1260
	atcttgaatc	tt					1272
15	<210> 84						
	<211> 7						
	<212> DNA						
	<213> Arti1	ficial Seque	ence				
20	<220>			_			
	<223> Desci	ription of A	Artificial :	Sequence: S	Synthetic		
•	<400> 84						
	<400> 04						
	~~~~						7
25	gcggaag						7 .
25							7
25	<210> 85						7
25	<210> 85 <211> 1463					÷	7
25	<210> 85 <211> 1463 <212> DNA	sapiens					7 .
	<210> 85 <211> 1463	sapiens					<b>7</b> .
25	<210> 85 <211> 1463 <212> DNA	sapiens					7
	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85	•	atgtgtacct	gecetggge	aaggtgggcc	tgacttgtta	
	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattetgta	, agcatttcct				tgacttgtta gcccagcatg	60
	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattctgta gagtgttaga	, agcatttcct gttttaccct	gttcctctag	gagggcctgg	taccaccaca		60
	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattctgta gagtgttaga gtgtggtgcc	agcatttect gttttaccet tcagcaggag	gttcctctag gcatctggtt	gagggcctgg acaatcaaca	taccaccaca caagctgttc	gcccagcatg	60 120 180
30	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattetgta gagtgttaga gtgtggtgcc aaagaaactt	agcatttect gttttaccet tcagcaggag caggaggaat	gttcctctag gcatctggtt agggttttag	gagggcctgg acaatcaaca gagggcatgg	taccaccaca caagctgttc ggaccctcct	gcccagcatg cagccaattt	60 120 180 240
30	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattctgta gagtgttaga gtgtggtgcc aaagaaactt ccaggatgtg	agcatttcct gttttaccct tcagcaggag caggaggaat ccaccaatca	gttcctctag gcatctggtt agggttttag taaggaggca	gagggcctgg acaatcaaca gagggcatgg ggggcctcct	taccaccaca caagetgtte ggacceteet teegetgete	gcccagcatg cagccaattt gcacccgaag	60 120 180 240 300
30	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattetgta gagtgttaga gtgtggtgcc aaagaaactt ccaggatgtg tctaggtgtc	agcatttcct gttttaccct tcagcaggag caggaggaat ccaccaatca cgtggcctca	gttcctctag gcatctggtt agggttttag taaggaggca gccccctct	gagggcctgg acaatcaaca gagggcatgg ggggcctcct gcacacctgc	taccaccaca caagetgtte ggaccetect teegetgete atetteette	gcccagcatg cagccaattt gcacccgaag cctgggactc	60 120 180 240 300 360
30	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattctgta gagtgttaga gtgtggtgcc aaagaaactt ccaggatgtg tctaggtgtc cctctgctt	agcatttcct gttttaccct tcagcaggag caggaggaat ccaccaatca cgtggcctca aagcgtaaac	gttcctctag gcatctggtt agggttttag taaggaggca gccccctct atggatgccc	gagggcctgg acaatcaaca gagggcatgg ggggcctcct gcacacctgc aggacctggc	taccaccaca caagctgttc ggaccctcct tccgctgctc atcttccttc ctcaatcttc	gcccagcatg cagccaattt gcacccgaag cctgggactc tcatcagctt	60 120 180 240 300 360 420
30	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattetgta gagtgttaga gtgtggtgcc aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg	agcattteet gttttaeeet teagcaggag caggaggaat ecaceaatea egtggeetea aagegtaaae taetgaeagt	gttcctctag gcatctggtt agggttttag taaggaggca gccccctct atggatgccc gtgagaccct	gagggcctgg acaatcaaca gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga	taccaccaca caagetgtte ggaccetect teegetgete atetteette etcaatette teaatecect	gcccagcatg cagccaattt gcacccgaag cctgggactc tcatcagctt cgagtctggt	60 120 180 240 300 360 420 480
30	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattetgta gagtgttaga gtgtggtgcc aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg cttccctgtg	agcatttect gttttaccct tcagcaggag caggaggaat ccaccaatca cgtggcctca aagcgtaaac tactgacagt caatcaatgg	gttcctctag gcatctggtt agggttttag taaggaggca gccccctct atggatgccc gtgagaccct aagccagcga	gagggcctgg acaatcaaca gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga ggcagggtca	taccaccaca caagetgtte ggacceteet teegetgete atetteette etcaatette teaateceet catgeeeegt	gcccagcatg cagccaattt gcacccgaag cctgggactc tcatcagctt cgagtctggt gggttggtga	60 120 180 240 300 360 420 480 540
30	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattctgta gagtgttaga gtgtggtgcc aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg cttccctgtg agacttggag	agcatttcct gttttaccct tcagcaggag caggaggaat ccaccaatca cgtggcctca aagcgtaaac tactgacagt caatcaatgg aaggaacgtg	gttcctctag gcatctggtt agggttttag taaggaggca gccccctct atggatgccc gtgagaccct aagccagcga ggcaagtctt	gagggcctgg acaatcaaca gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga ggcagggtca cccaggaaca	taccaccaca caagctgttc ggaccctcct tccgctgctc atcttccttc ctcaatcttc tcaatcccct catgccccgt ggtagggcag	gcccagcatg cagccaattt gcacccgaag cctgggactc tcatcagctt cgagtctggt gggttggtga ttagaggtgc	60 120 180 240 300 360 420 480 540 600
30	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattetgta gagtgttaga gtgtggtgcc aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc	agcattteet gttttaecet teagcaggag caggaggaat ecaceaatea egtggeetea aagegtaaae taetgaeagt caateaatgg aaggaaegtg tggtgeagee	gttcctctag gcatctggtt agggttttag taaggaggca gccccctct atggatgccc gtgagaccct aagccagcga ggcaagtctt cggttcggag	gagggcctgg acaatcaaca gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga ggcagggtca cccaggaaca caggaagacg	taccaccaca caagetgtte ggacceteet teegetgete atetteette ctcaatette teaatecet catgecegt ggtagggeag cttaataaat	gcccagcatg cagccaattt gcacccgaag cctgggactc tcatcagctt cgagtctggt gggttggtga ttagaggtgc ggaggaaagg	60 120 180 240 300 360 420 480 540 600 660
30	<210> 85 <211> 1463 <212> DNA <213> Homo <400> 85 gaattetgta gagtgttaga gtgtggtgcc aaagaaactt ccaggatgtg tctaggtgtc cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc tgcaggacac	agcattteet gttttaeeet teageaggag caggaggaat ceaceaatea egtggeetea aagegtaaae taetgaeagt caateaatgg aaggaaegtg tggtgeagee aggeaaaggt	gttcctctag gcatctggtt agggttttag taaggaggca gccccctct atggatgccc gtgagaccct aagccagcga ggcaagtctt cggttcggag gctgagctgg	gagggcctgg acaatcaaca gagggcatgg ggggcctcct gcacacctgc aggacctggc actcctctga ggcagggtca cccaggaaca caggaagacg accctttatt	taccaccaca caagctgttc ggaccctcct tccgctgctc atcttccttc ctcaatcttc tcaatccct catgccccgt ggtagggcag cttaataaat tctgcccttc	gcccagcatg cagccaattt gcacccgaag cctgggactc tcatcagctt cgagtctggt gggttggtga ttagaggtgc ggaggaaagg gctgatagac	60 120 180 240 300 360 420 480 540 600 660 720

	agctcagtga	atctgggcga	gtaacacaaa	acttgagtgt	ccttacctga	aaaatagagg	900
	•		attgtgtgtg			•	
			gtggagccca				
	-		ttttatgtat				
5			tggcacataa				
			tattacctgt				
	=		ctgaaaaaca				
			catctgtcaa				
			aggtgggctt				
10			tttgtctatg				
		tggttatgga		3333333			1463
		0550000550					
	<210> 86					•	
	<211> 83						
15	<212> DNA						
	<213> Homo	sapiens		•			
	12237 110	04p-0115					
	<400> 86						
		agtgccactg	tttgtctatg	gagaggagg	cctcagtgct	qaqqqccaaq	60
20		tggttatgga		3 3 333 33	3 <b>3</b>	3 333 3	83
	<b>3</b>	- 5555					
	<210> 87	•					
	<211> 614						
	<212> DNA						
25	<213> Homo	sapiens					
		-					
	<400> 87						
	gaattctgta	agcatttcct	atgtgtacct	gcccctgggc	aaggtgggcc	tgacttgtta	60
	gagtgttaga	gttttaccct	gttcctctag	gagggcctgg	taccaccaca	gcccagcatg	120
30	gtgtggtgcc	tcagcaggag	gcatctggtt	acaatcaaca	caagctgttc	cagccaattt	180
	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	gcacccgaag	240
	ccaggatgtg	ccaccaatca	taaggaggca	ggggcctcct	tccgctgctc	cctgggactc	3 0 0
	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	360
	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	420
35	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	480
	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	540
	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
	ggggcatctc	tggt					614
-							
40	<210> 88						
	<211> 661						
	<212> DNA				•		
,	<213> Homo	sapiens					

	<400> 88						
	gaattctgta	agcatttcct	atgtgtacct	gcccctgggc	aaggtgggcc	tgacttgtta	60
					taccaccaca		
	gtgtggtgcc	tcagcaggag	gcatctggtt	acaatcaaca	caagctgttc	cagccaattt	180
5	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	gcacccgaag	240
	ccaggatgtg	ccaccaatca	taaggaggca	ggggcctcct	tccgctgctc	cctgggactc	300
	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	360
	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	420
	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	480
10	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	540
	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
	ggggcatctc	tggtgcagcc	cggttcggag	caggaagacg	cttaataaat	gctgatagac	660
	t						661
15	<210> 89						
	<211> 7	•					
	<212> DNA						
	<213> Homo	sapiens					
					•		
20	<400> 89						
	gaggaaa						7
					•		
	<210> 90	•					
25	<211> 7						
23	<212> DNA						
	<213> Homo	sapiens					
	<400> 90						
	caggaag					٠	7
30	0499449						•
	<210> 91						
	<211> 102				•		
	<212> DNA						
	<213> Homo	sapiens					
35		_					
	<400> 91	•					
	tattttatat	atataatata	tatataaaat	atataatata	caatataaat	atatagtgtg	60
		cgtgtgtgta					102
	_	· - <del>-</del>	_				
40	<210> 92						
	<211> 1380						
	<212> DNA						
	<213> Homo	caniene					

	<400> 92						
	gaattctgta	agcatttcct	atgtgtacct	gcccctgggc	aaggtgggcc	tgacttgtta	60
	gagtgttaga	gttttaccct	gttcctctag	gagggcctgg	taccaccaca	gcccagcatg	120
	gtgtggtgcc	tcagcaggag	gcatctggtt	acaatcaaca	caagctgttc	cagccaattt	180
5	aaagaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	gcacccgaag	240
	ccaggatgtg	ccaccaatca	taaggaggca	ggggcctcct	tccgctgctc	cctgggactc	300
•	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	360
	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	420
	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	480
10	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	540
	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	600
	ggggcatctc	tggtgcagcc	cggttcggag	caggaagacg	cttaataaat	gctgatagac	660
	tgcaggacac	aggcaaaggt	gctgagctgg	accetttatt	tctgcccttc	tcccttctgg	720
	caccccggcc	aggaaattgc	tgcagccttt	ctggaatccc	gttcatttt	cttactggtc	780
15	cacaaaaggg	gccaaatgga	agcagcaaga	cctgagttca	aattaaatct	gccaactacc	840
	agctcagtga	atctgggcga	gtaacacaaa	acttgagtgt	ccttacctga	aaaatagagg	900
	ttagagggat	gctatgtgcc	attgtgtgtg	tgtgttgggg	gtggggattg	ggggtgattt	960
	gtgagcaatt	ggaggtgagg	gtggagccca	gtgcccagca	cctatgcact	ggggacccaa	1020
	aaaggagcat	cttctcatga	ttttatgtat	cagaaattgg	gatggcatgt	cattgggaca	1080
20	gcgtcttttt	tcttgtatgg	tggcacataa	atacatgtgt	cttataatta	atggtatttt	1140
	agatttgacg	aaatatggaa	tattacctgt	tgtgctgatc	ttgggcaaac	tataatatct	1200
	ctgggcaaaa	atgtccccat	ctgaaaaaca	gggacaacgt	tcctccctca	gccagccact	1260
	atggggctaa	aatgagacca	catctgtcaa	gggttttgcc	ctcacctccc	tecetgetgg	1320
	atggcatcct	tggtaggcag	aggtgggctt	cgggcagaac	aagccgtgct	gagctaggac	1380
25							
	<210> 93						
	<211> 154						
	<212> DNA					•	
••	<213> Homo	sapiens			•		
30							
	<400> 93						
				gttattttat			
				tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagcaataag	ccat	•		154
35							
	<210> 94	•					
	<211> 153						
	<212> DNA						
40	<213> Homo	sapiens					
40							
	<400> 94						
				ttattttata			
				gtgtgtgtat	gcgtgtgtgt	agacacacac	
45	gcatacacac	atataatgga	agcaataagc	cat			153
45							

	<210> 95						
•	<211> 152						
	<212> DNA						
	<213> Homo	sapiens					
5							
	<400> 95						
	gggggaaaag	tttctttcag	agagttaagt	tattttatat	atataatata	tatataaaat	60
	atataatata	caatataaat	atatagtgtg	tgtgtgtatg	cgtgtgtgta	gacacacacg	120
	catacacaca	tataatggaa	gcaataaġcc	at			152
10.			•				
	<210> 96						
	<211> 151	•					
	<212> DNA						
	<213> Homo	sapiens					•
15							
	<400> 96						
	ggggaaaagt	ttctttcaga	gagttaagtt	attttatata	tataatatat	atataaaata	60
	tataatatac	aatataaata	tatagtgtgt	gtgtgtatgc	gtgtgtgtag	acacacacgc	120
	atacacacat	ataatggaag	caataagcca	t			151
20							
	<210> 97						
	<211> 150						
	<212> DNA						•
	<213> Homo	sapiens					
25							
	<400> 97						
	gggaaaagtt	tctttcagag	agttaagtta	ttttatatat	ataatatata	tataaaatat	60
٠.	atatataca	atataaatat	atagtgtgtg	tgtgtatgcg	tgtgtgtaga	cacacacgca	120
	tacacacata	taatggaagc	aataagccat				150
30						-	
	<210> 98						
	<211> 149						
	<212> DNA						
	<213> Homo	sapiens					
35							
	<400> 98					•	
	ggaaaagttt	ctttcagaga	gttaagttat	tttatatata	taatatatat	ataaaatata	60
	taatatacaa	tataaatata	tagtgtgtgt	gtgtatgcgt	gtgtgtagac	acacacgcat	120
	acacacatat	aatggaagca	ataagccat				149
40							
	<210> 99	•	•				
	<211> 148						
	<212> DNA						
4.5	<213> Homo	sapiens					
45							•

	<400> 99					•	
	gaaaagtttc	tttcagagag	ttaagttatt	ttatatatat	aatatatata	taaaatatat	60 ·
	aatatacaat	ataaatatat	agtgtgtgtg	tgtatgcgtg	tgtgtagaca	cacacgcata	120
	cacacatata	atggaagcaa	taagccat				148
5							
	<210> 100						
	<211> 147						
	<212> DNA			•			
	<213> Homo	sapiens			•		
10							
	<400> 100						
	aaaagtttct	ttcagagagt	taagttattt	tatatatata	atatatatat	aaaatatata	60
	atatacaata	taaatatata	gtgtgtgtgt	gtatgcgtgt	gtgtagacac	acacgcatac	120
	acacatataa	tggaagcaat	aagccat				147
15							
	<210> 101						
	<211> 146	•				-	
	<212> DNA						
	<213> Homo	sapiens					
20							
	<400> 101			•			
	aaagtttctt	tcagagagtt	aagttatttt	atatatata	tatatata	aaatatataa	60
	tatacaatat	aaatatatag	tgtgtgtg	tatgcgtgtg	tgtagacaca	cacgcataca	120
	cacatataat	ggaagcaata	agccat				146
25							
	<210> 102						
	<211> 145		•				
	<212> DNA						
	<213> Homo	sapiens					
30							•
	<400> 102					•	
	aagtttcttt	cagagagtta	agttattta	tatatataat	atatataa	aatatataat	60
	atacaatata	aatatatagt	gtgtgtgtgt	atgcgtgtgt	gtagacacac	acgcatacac	120
	acatataatg	gaagcaataa	gccat				145
35							
	<210> 103						
	<211> 144						
	<212> DNA						
40	<213> Homo	sapiens					
40							
	<400> 103						
		agagagttaa		•			
		atatatagtg		tgcgtgtgtg	tagacacaca	cgcatacaca	
15	catataatgg	aagcaataag	ccat				144
<b>4</b> 5							

	<210> 104						
	<211> 143	•					
	<212> DNA						
	<213> Homo	sapiens					
5							
	<400> 104						
	gtttcttca	gagagttaag	ttattttata	tatataatat	atatataaaa	tatataatat	60
	acaatataaa	tatatagtgt	gtgtgtgtat	gcgtgtgtgt	agacacacac	gcatacacac	120
	atataatgga	agcaataagc	cat				143
10					•		
	<210> 105						
	<211> 142						
	<212> DNA						
	<213> Homo	sapiens					
15							
	<400> 105						,
	tttctttcag	agagttaagt	tattttatat	atataatata	tatataaaat	atataatata	€0
	caatataaat	atatagtgtg	tgtgtgtatg	cgtgtgtgta	gacacacacg	catacacaca	120
	tataatggaa	gcaataagcc	at				142
20							
	<210> 106						
	<211> 141	_					
	<212> DNA	·					•
	<213> Homo	sapiens					
25							
	<400> 106	•					
	ttctttcaga	gagttaagtt	attttatata	tataatatat	atataaaata	tataatatac	60
	aatataaata	tatagtgtgt	gtgtgtatgc	gtgtgtgtag	acacacacgc	atacacacat	120
	ataatggaag	caataagcca	t				141
30				-	•		
	<210> 107						
	<211> 140				•		
	<212> DNA		•	•			
	<213> Homo	sapiens					
35							
	<400> 107						
		_				ataatataca	
			tgtgtatgcg	tgtgtgtaga	cacacacgca	tacacacata	120
	taatggaagc	aataagccat		-			140
40						•	
-	<210> 108						
	<211> 139						
	<212> DNA						
45	<213> Homo	sapiens	_		•		
45		•	•				

	<400> 108	•					
	ctttcagaga	gttaagttat	tttatatata	taatatatat	ataaaatata	taatatacaa	60
	tataaatata	tagtgtgtgt	gtgtatgcgt	gtgtgtagac	acacacgcat	acacacatat	120
	aatggaagca						139
5							
	<210> 109						
	<211> 138						
	<212> DNA						
	<213> Homo	sapiens					
10	•	_					
	<400> 109						
	tttcagagag	ttaagttatt	ttatatatat	aatatatata	taaaatatat	aatatacaat	60
					cacacgcata		
	atggaagcaa		3 3 3, 3	3.3.	, <b>.</b>		138
15		3					
	<210> 110						
	<211> 137						
	<212> DNA		•				
	<213> Homo	sapiens					
20							
	<400> 110	•					
	ttcagagagt	taagttattt	tatatata	atatatatat	aaaatatata	atatacaata	60
	taaatatata	gtgtgtgtgt	gtatgcgtgt	gtgtagacac	acacgcatac	acacatataa	120
	tggaagcaat				_		137
<b>25</b> .					•		
	<210> 111						
	<211> 136					•	
	<212> DNA						
	<213> Homo	sapiens					
30							
	<400> 111						•
	tcagagagtt	aagttatttt	atatataa	tatatata	aaatatataa	tatacaatat	60
	aaatatatag	tgtgtgtgtg	tatgcgtgtg	tgtagacaca	cacgcataca	cacatataat	120
	ggaagcaata	agccat					136
35		•					
	<210> 112						
	<211> 135						
	<212> DNA						
	<213> Homo	sapiens					
40							•
	<400> 112						
	cagagagtta	agttatttta	tatatataat	atatataa	aatatataat	atacaatata	60
	aatatatagt	gtgtgtgtgt	atgegtgtgt	gtagacacac	acgcatacac	acatataatg	120
	gaagcaataa	gccat					135
45							

WO 00/75279 PCT/US00/15728 <210> 113 <211> 134 <212> DNA <213> Homo sapiens 5 <400> 113 atatatagtg tgtgtgtgta tgcgtgtgtg tagacacaca cgcatacaca catataatgg 120 aagcaataag ccat 134 10 <210> 114 <211> 133 <212> DNA <213> Homo sapiens 15 <400> 114 tatatagtgt gtgtgtgtat gcgtgtgtgt agacacacac gcatacacac atataatgga 120 agcaataagc cat 133 20 <210> 115 <211> 132 <212> DNA <213> Homo sapiens 25 <400> 115 

30

. 35

atatagtgtg tgtgtgtatg cgtgtgtgta gacacacag catacacaca tataatggaa 120 gcaataagcc at 132

<210> 116

<211> 131 <212> DNA

<213> Homo sapiens

<400> 116

tatagtgtgt gtgtgtatgc gtgtgtgtag acacacacgc atacacacat ataatggaag 120 caataagcca t 131

<210> 117 <211> 130 <212> DNA

<213> Homo sapiens

45

40

	<400> 117						
	agttaagtta	ttttatatat	ataatatata	tataaaatat	ataatataca	atataaatat	60
	atagtgtgtg	tgtgtatgcg	tgtgtgtaga	cacacacgca	tacacacata	taatggaagc	120
	aataagccat						130
5							
	<210> 118						
	<211> 129						
	<212> DNA						
	<213> Homo	sapiens					
10							
	<400> 118						
	gttaagttat	tttatatata	taatatatat	ataaaatata	taatatacaa	tataaatata	60
						aatggaagca	
	ataagccat						129
15	_						
	<210> 119						
	<211> 128						
	<212> DNA						
	<213> Homo	sapiens					
20							
	<400> 119						
	ttaagttatt	ttatatatat	aatatatata	taaaatatat	aatatacaat	ataaatatat	60
	agtgtgtgtg	tgtatgcgtg	tgtgtagaca	cacacgcata	cacacatata	atggaagcaa	120
	taagccat						128
25					•		
	<210> 120						
	<211> 127						
	<212> DNA					•	
	<213> Homo	sapiens					
30		•					
	<400> 120		•				
	taagttattt	tatatata	atatatatat	aaaatatata	atatacaata	taaatatata	60
	gtgtgtgtgt	gtatgcgtgt	gtgtagacac	acacgcatac	acacatataa	tggaagcaat	120
	aagccat			•			127
35						•	
	<210> 121						
	<211> 126						
	<212> DNA						
	<213> Homo	sapiens					-
40		•					
	<400> 121						
	aagttatttt	atatataa	tatatata	aaatatataa	tatacaatat	aaatatatag	60
	tgtgtgtg	tatgcgtgtg	tgtagacaca	cacgcataca	cacatataat	ggaagcaata	120
	agccat						126
45		•					

WO 00/75279 PCT/US00/15728 <210> 122 <211> 125 <212> DNA <213> Homo sapiens 5 <400> 122 agttatttta tatatataat atatatataa aatatataat atacaatata aatatatagt 60

gtgtgtgtgt atgcgtgtgt gtagacacac acgcatacac acatataatg gaagcaataa 120 gccat 125

10 <210> 123 <211> 124 <212> DNA

<213> Homo sapiens

<213> Homo sapiens

15

20

25

30

35

40

45

<400> 123 tgtgtgtgta tgcgtgtgtg tagacacaca cgcatacaca catataatgg aagcaataag 120 ccat 124

<210> 124 <211> 123 <212> DNA

<400> 124 gtgtgtgtat gcgtgtgtgt agacacacac gcatacacac atataatgga agcaataagc 120

cat 123 <210> 125 <211> 153

<212> DNA <213> Homo sapiens

<400> 125 atatataata tacaatataa atatatagtg tgtgtgtgta tgcgtgtgtg tagacacaca 120 cgcatacaca catataatgg aagcaataag cca 153

<210> 126 <211> T52 <212> DNA <213> Homo sapiens

	<400> 126	•				•	
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagcaataag	cc			152
5							
	<210> 127						
	<211> 151						
	<212> DNA						
	<213> Homo	sapiens			•		
10						•	
	<400> 127		•				
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
		catataatgg				_	151
15							
	<210> 128						•
	<211> 150						•
	<212> DNA						
•	<213> Homo	sapiens				,	
20			•				
	<400> 128						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagcaataag				150
25							
	<210> 129						
	<211> 149						
	<212> DNA					-	
	<213> Homo	sapiens					
30							
	<400> 129						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagcaataa				149
35							
	<210> 130	•					
	<211> 148						
	<212> DNA		•				
	<213> Homo	sapiens					
40							
	<400> 130						
		_		=		tatatataaa	
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	12.0
	cgcatacaca	catataatgg	aagcaata				148
45							

	<210> 131						
	<211> 147						
	<212> DNA						
	<213> Homo	sapiens.					
5							
	<400> 131	•					
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagcaat				147
10							
	<210> 132						
	<211> 146						
	<212> DNA						
	<213> Homo	sapiens	•		•		
13							
	<400> 132						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	1.20
	cgcatacaca	catataatgg	aagcaa				146
20				•			
	<210> 133						
	<211> 145						
	<212> DNA						
	<213> Homo	sapiens					
25							
	<400> 133						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aagca				145
30							
	<210> 134						
	<211> 144						
•	<212> DNA						
<b>.</b> .	<213> Homo	sapiens					
35		•			•		
	<400> 134						
		_	agagagttaa	_			
			atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	
40	cgcatacaca	catataatgg	aagc				144
40	_ :						
	<210> 135						
	<211> 143	•					
	<212> DNA						
15	<213> Homo	sapiens	٠.				

	<400> 135						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aag				143
5							
	<210> 136	•					
	<211> 142			•			
	<212> DNA		•				
	<213> Homo	sapiens					
10							
	<400> 136						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg	aa		•		142
15							
	<210> 137						
	<211> 141	·					
	<212> DNA						
	<213> Homo	sapiens	·				
20							
	<400> 137						
	ttgggggaaa	agtttcttţc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
•	cgcatacaca	catataatgg	a				141
25							
	<210> 138			-			
	<211> 140						
	<212> DNA						
	<213> Homo	sapiens					
30		•					•
	<400> 138						
		_		_	atatataata		
			atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataatgg					140
35						•	
	<210> 139						
	<211> 139						•
	<212> DNA						
40	<213> Homo	sapiens					
40							
	<400> 139						
					atatataata		
			atatatagtg	tgtgtgtgta	tgcgtgtåtg	tagacacaca	
15	cgcatacaca	catataatg					139
45							

	<210> 140						
	<211> 138					•	
	<212> DNA						
	<213> Homo	sapiens					
5		•					
	<400> 140						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataat					138
10							
	<210> 141						
	<211> 137					•	
	<212> DNA						
	<213> Homo	sapiens					
15							
	<400> 141						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
	cgcatacaca	catataa					137
20							
	<210> 142						
	<211> 136			•			
	<212> DNA						
	<213> Homo	sapiens					
25 .							
	<400> 142						
	ttgggggaaa	agtttctttc	agagagttaa	gttattttat	atatataata	tatatataaa	60
	atatataata	tacaatataa	atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	120
• •	cgcatacaca	catata					136
30		•					
	<210> 143					•	
	<211> 135	·					
	<212> DNA						
25	<213> Homo	sapiens			•		
35		,					
	<400> 143		•				
				gttattttat			
			atatatagtg	tgtgtgtgta	tgcgtgtgtg	tagacacaca	
40	cgcatacaca	catat					135
40	-210- 344						
	<210> 144						
	<211> 134						
	<212> DNA	annione			•		
45	<213> Homo	sapiens			·		

```
<400> 144
      atatataata tacaatataa atatatagtg tgtgtgtgta tgcgtgtgtg tagacacaca 120
      cgcatacaca cata
                                                                       134
 5
      <210> 145
      <211> 659
      <212> DNA
      <213> Homo sapiens
10
      <400> 145
      gaattetgta agcattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
15
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctect tecgetgete eetgggacte 300
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccet actectetga teaateceet gggttggtga 480
20
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat gctgataga 659
      <210> 146
25
      <211> 658
      <212> DNA
      <213> Homo sapiens
      <400> 146
30
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatq 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
35
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
40
      ggggcatete tggtgcagee eggtteggag caggaagaeg ettaataaat getgatag
      <210> 147
      <211> 657
      <212> DNA
45
      <213> Homo sapiens
```

```
<400> 147
      gaattetgta agcattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
 5
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgto egtggcotca geocecetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
10
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat gctgata
      <210> 148
15
      <211> 656
      <212> DNA
      <213> Homo sapiens
      <400> 148
20
      gaattetgta agcattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
25
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540°
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
30
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat gctgat
                                                                         656
      <210> 149
      <211> 655
      <212> DNA
35
      <213> Homo sapiens
      <400> 149
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
40
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
45
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
```

```
cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtett cecaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee cggtteggag caggaagaeg ettaataaat getga
 5
      <210> 150
      <211> 654.
      <212> DNA
      <213> Homo sapiens
10
      <400> 150
      gaattetgta agcattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccaqcatq 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
15
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gcccccctct gcacacctgc atcttccttc tcatcaqctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
20
      agacttggag aaggaacgtg ggcaagtett cecaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat gctg
      <210> 151
      <211> 653
25
      <212> DNA
      <213> Homo sapiens
      <400> 151
      gaattetgta agcattteet atgtgtaeet geeectggge aaggtgggee tgaettgtta 60
30
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
35
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat gct
40
      <210> 152*
      <211> 652
      <212> DNA
      <213> Homo sapiens
```

45

```
<400> 152
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
 5
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccqaaq 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgqqactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
10
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee eggtteggag caggaagaeg ettaataaat qe
      <210> 153
15
      <211> 651
      <212> DNA
      <213> Homo sapiens
      <400> 153
20
      gaattetgta ageattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatq 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
25
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
30
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataaat g
                                                                         651
      <210> 154
      <211> 650
      <212> DNA
35
      <213> Homo sapiens
      <400> 154
      gaattetgta ageattteet atgtgtacet geeectggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca qcccaqcatq 120
40
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgto cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
45
      acttatggtg tactgacagt gtgagaccet acteetetga teaateeeet gggttggtga 480
```

```
cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee eggtteggag caggaagaeg ettaataaat
5
      <210> 155
      <211> 649
      <212> DNA
      <213> Homo sapiens
10
      <400> 155
      gaattetgta agcattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
15
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
20
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatete tggtgcagee eggtteggag caggaagaeg ettaataaa
      <210> 156
      <211> 648
25
      <212> DNA
      <213> Homo sapiens
      <400> 156
      gaattetgta ageattteet atgtgtaeet geeectggge aaggtgggee tgaettgtta 60
30
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
35
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaataa
40
      <210> 157
      <211> 647
      <212> DNA
      <213> Homo sapiens
45
```

```
<400> 157
      gaattetgta ageattteet atgtgtacet geeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca qcccaqcatq 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
 5
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaaq 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgto cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagegtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
10
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaata
                                                                         647
      <210> 158
15
      <211> 646
      <212> DNA
      <213> Homo sapiens
      <400> 158
20
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
25
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480 .
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
30
      ggggcatete tggtgcagee eggtteggag eaggaagaeg ettaat
                                                                         646
      <210> 159
      <211> 645
      <212> DNA
35
      <213> Homo sapiens
      <400> 159
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
40
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccqaaq 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
45
      acttatggtg tactgacagt gtgagaccet acteetetga teaateeeet gggttggtga 480°
```

```
cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
       ggggcatctc tggtgcagcc cggttcggag caggaagacg cttaa
 5
       <210> 160
       <211> 644
       <212> DNA
       <213> Homo sapiens
10
       <400> 160
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
15
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cetetgettt aagegtaaac atggatgeec aggaeetgge etcaatette egagtetggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
20
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg ctta
                                                                         644
      <210> 161
      <211> 643
25
      <212> DNA
      <213> Homo sapiens
      <400> 161
      gaattetgta agcattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
30
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagett 360
35
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cqaqtctqqt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg ctt
                                                                         643
40
      <210> 162
      <211> 642
      <212> DNA
      <213> Homo sapiens
45
```

```
<400> 162
      gaattetgta agcattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
 5
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
10
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaagacg ct
                                                                         642
      <210> 163
15
      <211> 641
      <212> DNA
      <213> Homo sapiens
      <400> 163
20
      gaattetgta agcattteet atgtgtacet geeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
25
      totaggtgto egtggcotca geoccectot geacacetge atotteette teatcagett 360
      cetetgettt aagegtaaac atggatgeee aggaeetgge eteaatette egagtetggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
30
      ggggcatete tggtgcagee eggtteggag eaggaagaeg e
                                                                         641
      <210> 164
      <211> 640
      <212> DNA
35
      <213> Homo sapiens
      <400> 164
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
40
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgto ogtggcotca gooccootct goacacotgo atottootto toatcagott 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
45
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
```

```
cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
       ggggcatete tggtgcagee eggtteggag caggaagaeg
 5
       <210> 165
       <211> 639
       <212> DNA
       <213> Homo sapiens
10
       <400> 165
       gaattctgta agcatttcct atgtgtacct gcccctgggc aaggtgggcc tgacttgtta 60
       gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
       gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
       aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
15
       ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
       tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
       cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
       acttatggtg tactgacayt gtgagaccct actcctctga tcaatcccct gggttggtga 480
       cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtqc 540
20
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
       ggggcatete tggtgcagee eggtteggag eaggaagae
       <210> 166
       <211> 638
25
       <212> DNA
       <213> Homo sapiens
       <400> 166
       gaattetgta agcattteet atgtgtacet geeettggge aaggtgggee tgaettgtta 60
30
       gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
       gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
       aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
       ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
       tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
35
       cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
       acttatggtg tactgacayt gtyagaccct actcctctga tcaatcccct gggttggtga 480
       cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
       ggggcatctc tggtgcagcc cggttcggag caggaaga
                                                                          638
40
       <210> 167
       <211> 637
       <212> DNA
       <213> Homo sapiens
```

45

```
<400> 167
      gaattetgta ageattteet atgtgtacet geceetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
 5
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggtgte egtggeetea geeceetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
10
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tggtgcagcc cggttcggag caggaag
                                                                         637
      <210> 168
15
      <211> 650
      <212> DNA
      <213> Homo sapiens
      <400> 168
20
      agcatttcct atgtgtacct gcccctgggc aaggtgggcc tgacttgtta gagtgttaga 60
      gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg gtgtggtgcc 120
      tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt aaagaaactt 180
      caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag ccaggatgtg 240
      ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc tctaggtgtc 300
25
      egtggeetea geceeetet geacacetge atetteette teateagett eetetgettt 360
      aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt acttatggtg 420
      tactgacagt gtgagaccct actectetga teaateceet gggttggtga ettecetgtg 480
      caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc agacttggag 540
      aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg ggggcatctc 600
30
      tggtgcagcc cggttcggag caggaagacg cttaataaat gctgatagac
                                                                         650
      <210> 169
      <211> 640
      <212> DNA
35
      <213> Homo sapiens
      <400> 169
      atgtgtacct gcccctgggc aaggtgggcc tgacttgtta gagtgttaga gttttaccct 60
      gttcctctag gagggcctgg taccaccaca gcccagcatg gtgtggtgcc tcagcaggag 120
40
      gcatctggtt acaatcaaca caagctgttc cagccaattt aaagaaactt caggaggaat 180
      agggttttag gagggcatgg ggaccetect geaccegaag ceaggatgtg ceaccaatea 240
      taaggaggca ggggcctcct tccgctgctc cctgggactc tctaggtgtc cgtggcctca 300
      gececetet geacacetge atetteette teateagett cetetgettt aagegtaaac 360
      atggatgccc aggacctggc ctcaatcttc cgagtctggt acttatggtg tactgacagt 420
45
```

gtgagaccct actcctctga tcaatcccct gggttggtga cttccctgtg caatcaatgg 480



	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	agacttggag	aaggaacgtg	540
	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaagg	ggggcatctc	tggtgcagcc	600
	cggttcggag	caggaagacg	cttaataaat	gctgatagac			640
_					•		
5	<210> 170						
	<211> 610						
	<212> DNA						
	<213> Homo	sapiens					
10							
10 -	<400> 170						
		gagtgttaga					
		gtgtggtgcc					
		aaagaaactt					
1.5		ccaggatgtg		•			
15		tctaggtgtc					
		cctctgcttt					
	•	acttatggtg					
		cttccctgtg					
•	ttagaggtgc	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	540
20		ggggcatctc	tggtgcagcc	cggttcggag	caggaagacg	cttaataaat	600
	gctgatagac						610
	<210> 171						
25	<211> 300						
25	<211> 300 <212> DNA						
25	<211> 300	sapiens				·	
25	<211> 300 <212> DNA <213> Homo	sapiens				·	
25	<211> 300 <212> DNA <213> Homo <400> 171				grantatia		60
	<211> 300 <212> DNA <213> Homo <400> 171 cctctgcttt	aagcgtaaac					
25	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg	aagcgtaaac tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	120
	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg	aagcgtaaac tactgacagt caatcaatgg	gtgagaccct aagccagcga	actectetga ggcagggtca	tcaatcccct catgccccgt	gggttggtga ttagaggtgc	120 180
	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag	aagcgtaaac tactgacagt caatcaatgg aaggaacgtg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtea cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	120 180 240
	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag	aagcgtaaac tactgacagt caatcaatgg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtea cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	120 180 240
30	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc	aagcgtaaac tactgacagt caatcaatgg aaggaacgtg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtea cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	120 180 240
	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc  <210> 172	aagcgtaaac tactgacagt caatcaatgg aaggaacgtg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtea cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	120 180 240
30	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc  <210> 172 <211> 100	aagcgtaaac tactgacagt caatcaatgg aaggaacgtg	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtea cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	120 180 240
30	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc  <210> 172 <211> 100 <212> DNA	aagcgtaaac tactgacagt caatcaatgg aaggaacgtg tggtgcagcc	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtea cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	120 180 240
30	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc  <210> 172 <211> 100	aagcgtaaac tactgacagt caatcaatgg aaggaacgtg tggtgcagcc	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtea cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	120 180 240
30 35	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc  <210> 172 <211> 100 <212> DNA <213> Homo	aagcgtaaac tactgacagt caatcaatgg aaggaacgtg tggtgcagcc	gtgagaccct aagccagcga ggcaagtctt	actectetga ggcagggtca cccaggaaca	tcaatcccct catgccccgt ggtagggcag	gggttggtga ttagaggtgc ggaggaaagg	120 180 240
30	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc  <210> 172 <211> 100 <212> DNA <213> Homo  <400> 172	aagcgtaaac tactgacagt caatcaatgg aaggaacgtg tggtgcagcc	gtgagaccct aagccagcga ggcaagtctt cggttcggag	actectetga ggcagggtca cccaggaaca caggaagacg	tcaatcccct catgccccgt ggtagggcag cttaataaat	gggttggtga ttagaggtgc ggaggaaagg gctgatagac	120 180 240 300
30 35	<211> 300 <212> DNA <213> Homo  <400> 171 cctctgcttt acttatggtg cttccctgtg agacttggag ggggcatctc  <210> 172 <211> 100 <212> DNA <213> Homo  <400> 172 ggcaagtctt	aagcgtaaac tactgacagt caatcaatgg aaggaacgtg tggtgcagcc	gtgagaccct aagccagcga ggcaagtctt cggttcggag	actcctctga ggcagggtca cccaggaaca caggaagacg	tcaatcccct catgccccgt ggtagggcag cttaataaat	gggttggtga ttagaggtgc ggaggaaagg gctgatagac	120 180 240 300

```
<210> 173
      <211> 71.
      <212> DNA
      <213> Homo sapiens
 5
      <400> 173
      gggaggaaag gggggcatct ctggtgcagc ccggttcgga gcaggaagac gcttaataaa 60
      tgctgataga c
                                                                         71
10
      <210> 174
      <211> 613
      <212> DNA
      <213> Homo sapiens
15
      <400> 174
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
20
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggtgte egtggeetea geeceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagegtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      ettecetgtg caatcaatgg aagceagega ggeagggtea catgeeeegt ttagaggtge 540
25
      agacttggag aaggaacgtg ggcaagtett cecaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tgg
                                                                         613
      <210> 175
      <211> 612
30
      <212> DNA
      <213> Homo sapiens
      <400> 175
      gaattetgta agcattteet atgtgtacet geeectggge aaggtgggee tgaettgtta 60
35
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
40
      cctctgcttt aagegtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc tg
                                                                         612
45
```

```
<210> 176
       <211> 611
       <212> DNA
       <213> Homo sapiens
 5
       <400> 176
       gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
       gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
       gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
10
       aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaaq 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
       tctaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cetetgettt aagegtaaac atggatgeec aggacetgge etcaatette egagtetggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
15
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatctc t
                                                                          611
      <210> 177
20
      <211> 610
       <212> DNA
      <213> Homo sapiens
      <400> 177
25
      gaattotyta agcatttoot atytytaeet geeeetygge aagytyggee tyaettyttá 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccaqcatq 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
30
      totaggtgto ogtggootoa goodooctot goadacetgo atottootto toatcagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      ettecetgtg caatcaatgg aagecagega ggeagggtea catgeceegt ttagaggtge 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
35
      ggggcatctc
                                                                         610
      <210> 178
      <211> 609
      <212> DNA
40
      <213> Homo sapiens
      <400> 178
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
45
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
```

```
aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cqtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
5
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggcatct
                                                                         609
10
      <210> 179
      <211> 608
      <212> DNA
      <213> Homo sapiens
15
      <400> 179
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
20
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
25
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
                                                                         608
      ggggcatc
      <210> 180
      <211> 607
30
      <212> DNA
      <213> Homo sapiens
      <400> 180
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
35
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
40
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
                                                                         607
      ggggcat
```

45

```
<210> 181
       <211> 606
       <212> DNA
      <213> Homo sapiens
 5
       <400> 181
      gaattetgta ageattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
10
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
15
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
      ggggca
                                                                         606
      <210> 182
20
      <211> 605
      <212> DNA
      <213> Homo sapiens
      <400> 182
25
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
30
      tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      ettecetgtg caatcaatgg aagecagega ggeagggtea catgeceegt ttagaggtge 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
35
      ggggc
                                                                         605
      <210> 183
      <211> 604
      <212> DNA
40
      <213> Homo sapiens
      <400> 183
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
45
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
```

	aaaqaaactt	caggaggaat	agggttttag	gagggcatgg	ggaccctcct	acacccaaa	240
		ccaccaatca					
		cgtggcctca					
		aagcgtaaac					
5		tactgacagt					
		caatcaatgg					
		aaggaacgtg					
	9999	aaggaacgeg	ggoudgeece	cccaggaaca	ggcagggcag	ggaggaaagg	604
	9999						604
10	<210> 184						
	<211> 603						
	<212> DNA						
	<213> Homo	sapiens					
	12132 1101110	Dupiens					
15	<400> 184					•	
		agcatttcct	atgtgtacct	accetaaac	aaggtgggg	taacttatta	60
		gttttaccct					
		tcagcaggag					
		caggaggaat	•				
20		ccaccaatca					
20							
		cgtggcctca					
		aagcgtaaac					
		tactgacagt					
25		caatcaatgg					
23		aaggaacgtg	ggcaagtett	cccaggaaca	ggcaggcag	ggaggaaagg	
	999	•					603
	<210> 185						
	<211> 602						
30	<212> DNA						
	<213> Homo	saniens					•
•	12137 1101110	Baptens					
	<400> 185						
		agcatttcct	atgtgtacct	accectagae	aaggtgggcc	tgacttgtta	60
35		gttttaccct					
		tcagcaggag					
		caggaggaat					
		ccaccaatca					
		cgtggcctca					
40		aagcgtaaac					
		tactgacagt					
		caatcaatgg					
		aaggaacgtg					
	99	aaggaacgcg	JJCAAJCCCC	cccayyaaca	ggragggrag	ggaggaaagg	602
	ככ						002

45

```
<210> 186
      <211> 601
      <212> DNA
      <213> Homo sapiens
5
      <400> 186
      gaattetgta agcattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
10
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420.
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
15
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
                                                                         601
      <210> 187
20
      <211> 600
      <212> DNA
      <213> Homo sapiens
      <400> 187
25
      gaattetgta ageattteet atgtgtacet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
30
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccet actcetetga teaateceet gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaagg 600
35
      <210> 188
      <211> 599
      <212> DNA
      <213> Homo sapiens
40
      <400> 188
      gaattetgta ageattteet atgtgtaeet geeeetggge aaggtgggee tgaettgtta 60
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
45
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
```

```
ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
       tetaggtgte egtggeetea geceeetet geacacetge atetteette teateagett 360
       cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctgqt 420
       acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
 5
       cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
       agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaag 599
       <210> 189
       <211> 598
10
       <212> DNA
       <213> Homo sapiens
       <400> 189
      gaattetgta agcattteet atgtgtacet geeettggge aaggtgggee tgaettgtta 60
15
      gagtgttaga gttttaccct gttcctctag gagggcctgg taccaccaca gcccagcatg 120
      gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca caagctgttc cagccaattt 180
      aaagaaactt caggaggaat agggttttag gagggcatgg ggaccctcct gcacccgaag 240
      ccaggatgtg ccaccaatca taaggaggca ggggcctcct tccgctgctc cctgggactc 300
      totaggtgtc cgtggcctca gccccctct gcacacctgc atcttccttc tcatcagctt 360
20
      cctctgcttt aagcgtaaac atggatgccc aggacctggc ctcaatcttc cgagtctggt 420
      acttatggtg tactgacagt gtgagaccct actcctctga tcaatcccct gggttggtga 480
      cttccctgtg caatcaatgg aagccagcga ggcagggtca catgccccgt ttagaggtgc 540
      agacttggag aaggaacgtg ggcaagtctt cccaggaaca ggtagggcag ggaggaaa
25
       <210> 190
      <211> 498
       <212> DNA
      <213> Homo sapiens
3.0
      <400> 190
      taccaccaca gcccagcatg gtgtggtgcc tcagcaggag gcatctggtt acaatcaaca 60
      caagetgtte cagecaattt aaagaaactt caggaggaat agggttttag gagggcatgg 120
      ggaccetect geaccegaag ceaggatgtg ceaceaatea taaggaggea ggggeeteet 180
      teegetgete cetgggaete tetaggtgte egtggeetea geececetet geacacetge 240
35
      atcttccttc tcatcagctt cctctgcttt aagcgtaaac atggatgccc aggacctggc 300
      ctcaatcttc cgagtctggt acttatggtg tactgacagt gtgagaccct actcctctga 360
      tcaatcccct gggttggtga cttccctgtg caatcaatgg aagccagcga ggcagggtca 420
      catgccccgt ttagaggtyc agacttggag aaggaacgtg ggcaagtctt cccaggaaca 480
      ggtagggcag ggaggaaa
                                                                         498
40
      <210> 191
      <211> 398
      <212> DNA
      <213> Homo sapiens
```

45

	<400> 191						
	agggttttag	gagggcatgg.	ggaccctcct	gcacccgaag	ccaggatgtg	ccaccaatca	60
	taaggaggca	ggggcctcct	teegetgete	cctgggactc	tctaggtgtc	cgtggcctca	120
	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	cctctgcttt	aagcgtaaac	180
5	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	acttatggtg	tactgacagt	240
•	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	cttccctgtg	caatcaatgg	300
	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	agacttggag	aaggaacgtg	360
	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaa			398
10	<210> 192						
٠	<211> 298						
	<212> DNA						
	<213> Homo	sapiens					
15	<400> 192						
	tctaggtgtc	cgtggcctca	gccccctct	gcacacctgc	atcttccttc	tcatcagctt	60
	cctctgcttt	aagcgtaaac	atggatgccc	aggacctggc	ctcaatcttc	cgagtctggt	120
	acttatggtg	tactgacagt	gtgagaccct	actcctctga	tcaatcccct	gggttggtga	180
	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	240
20	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaa	298
	<210> 193						
	<211> 198						
	<212> DNA						
25	<213> Homo	sapiens					
	<400> 193						
	ctcaatcttc	cgagtctggt	acttatggtg	tactgacagt	gtgagaccct	actcctctga	60
	tcaatcccct	gggttggtga	cttccctgtg	caatcaatgg	aagccagcga	ggcagggtca	120
30	catgccccgt	ttagaggtgc	agacttggag	aaggaacgtg	ggcaagtctt	cccaggaaca	180
	ggtagggcag	ggaggaaa					198
	<210> 194						•
	<211> 98						
35	<212> DNA						
	<213> Homo	sapiens					
	<400> 194						
	aagccagcga	ggcagggtca	catgccccgt	ttagaggtgc	agacttggag	aaggaacgtg	60
40	ggcaagtctt	cccaggaaca	ggtagggcag	ggaggaaa			98
	<210> 195						
	<211> 8	٠					
4.5	<212> DNA						
45	<213> HOMO	saniens					

<400> 195 8 ggaggaaa <210> 196 5 <211> 781 <212> DNA <213> Homo sapiens <400> 196 10 gggggcatct ctggtgcagc ccggttcgga gcaggaagac gcttaataaa tgctgataga 60 ctgcaggaca caggcaaagg tgctgagctg gaccctttat ttctgccctt ctcccttctg 120 gcaccccggc caggaaattg ctgcagcctt tctggaatcc cgttcatttt tcttactggt 180 ccacaaaagg ggccaaatgg aagcagcaag acctgagttc aaattaaatc tgccaactac 240 cageteagtg aatetgggeg agtaacacaa aaettgagtg teettaeetg aaaaatagag 300 15 gttagaggga tgctatgtgc cattgtgtgt gtgtgttggg ggtggggatt gggggtgatt 360 tgtgagcaat tggaggtgag ggtggagccc agtgcccagc acctatgcac tggggaccca 420 aaaaggagca tottotoatg attttatgta toagaaattg ggatggcatg toattgggac 480 agegtetttt ttettgtatg gtggeacata aatacatgtg tettataatt aatggtattt 540 tagatttgac gaaatatgga atattacctg ttgtgctgat cttgggcaaa ctataatatc 600 20 totgggcaaa aatgtoocca totgaaaaac agggacaacg ttootcootc agcoagcoac 660 tatggggcta aaatgagacc acatctgtca agggttttgc cctcacctcc ctccctgctg 720 gatggcatcc ttggtaggca gaggtgggct tcgggcagaa caagccgtgc tgagctagga 780 781 25 <210> 197 <211> 771 <212> DNA <213> Homo sapiens 30 <400> 197 ctggtgcagc ccggttcgga gcaggaagac gcttaataaa tgctgataga ctgcaggaca 60 caggcaaagg tgctgagctg gaccctttat ttctgccctt ctcccttctg gcaccccggc 120 caggaaattg ctgcagcctt tctggaatcc cgttcatttt tcttactggt ccacaaaagg 180 ggccaaatgg aagcagcaag acctgagttc aaattaaatc tgccaactac cagctcagtg 240 35 aatctgggcg agtaacacaa aacttgagtg teettaeetg aaaaatagaq qttaqaqqqa 300 tgctatgtgc cattgtgtgt gtgtgttggg ggtggggatt gggggtgatt tgtgagcaat 360 tggaggtgag ggtggagece agtgeecage acetatgeae tggggaecea aaaaggagea 420 tetteteatg attitatgta teagaaattg ggatggeatg teattgggae agegtetttt 480 ttcttgtatg gtggcacata aatacatgtg tcttataatt aatggtattt tagatttgac 540 40 gaaatatgga atattacctg ttgtgctgat cttgggcaaa ctataatatc tctgggcaaa 600

771

aatgtcccca tctgaaaaac agggacaacg ttcctccctc agccagccac tatggggcta 660 aaatgagacc acatctgtca agggttttgc cctcacctcc ctccctgctg gatggcatcc 720

ttggtaggca gaggtgggct tcgggcagaa caagccgtgc tgagctagga c

```
<210> 198
       <211> 761
       <212> DNA
       <213> Homo sapiens
 5
       <400> 198
      ccggttcgga gcaggaagac gcttaataaa tgctgataga ctgcaggaca caggcaaagg 60
      tgctgagctg gaccetttat ttetgeeett etecettetg geacceegge caggaaattg 120
      ctgcagcett tetggaatee egtteatttt tettaetggt ceacaaaagg ggccaaatgg 180
10
      aagcagcaag acctgagttc aaattaaatc tgccaactac cagctcagtg aatctgggcg 240
      agtaacacaa aacttgagtg teettaeetg aaaaatagag gttagaggga tgetatgtge 300
      cattgtgtgt gtgtgttggg ggtggggatt gggggtgatt tgtgagcaat tggaggtgag 360
      ggtggagccc agtgcccagc acctatgcac tggggaccca aaaaggagca tcttctcatg 420
      attttatgta tcagaaattg ggatggcatg tcattgggac agcgtctttt ttcttgtatg 480
15
      gtggcacata aatacatgtg tettataatt aatggtattt tagatttgac gaaatatgga 540
      atattacctg ttgtgctgat cttgggcaaa ctataatatc tctgggcaaa aatgtcccca 600
      tctgaaaaac agggacaacg ttcctccctc agccagccac tatggggcta aaatgagacc 660
      acatotytoa aggyttttyc cotoacotoo otocotycty yatyycatoo ttyytagyca 720
      gaggtgggct tcgggcagaa caagccgtgc tgagctagga c
                                                                         761
20
      <210> 199
      <211> 751
      <212> DNA
      <213> Homo sapiens
25
      <400> 199
      gcaggaagac gcttaataaa tgctgataga ctgcaggaca caggcaaagg tgctgagetg 60
      gaccetttat ttctgccctt ctcccttctg gcaccccggc caggaaattg ctgcagcctt 120
      tctggaatcc cgttcatttt tcttactggt ccacaaaagg ggccaaatgg aagcagcaag 180
30
      acctgagttc aaattaaatc tgccaactac cagctcagtg aatctgggcg agtaacacaa 240
      aacttgagtg teettaeetg aaaaatagag gttagaggga tgetatgtge eattgtgtgt 300
      gtgtgttggg ggtggggatt gggggtgatt tgtgagcaat tggaggtgag ggtggagccc 360
      agtgcccage acctatgcae tggggaccca aaaaggagca tetteteatg attttatgta 420
      tcagaaattg ggatggcatg tcattgggac agcgtctttt ttcttgtatg gtggcacata 480
35
      aatacatgtg tottataatt aatggtattt tagatttgac gaaatatgga atattacctg 540
      ttgtgctgat cttgggcaaa ctataatatc tctgggcaaa aatgtcccca tctgaaaaac 600
      agggacaacg ttcctccctc agccagccac tatggggcta aaatgagacc acatctgtca 660
      agggttttgc cctcacctcc ctccctgctg gatggcatcc ttggtaggca gaggtgggct 720
      tegggeagaa caageegtge tgagetagga e
                                                                         751
40
      <210> 200
      <211> 750
      <212> DNA
      <213> Homo sapiens
45
```

```
<400> 200
      caggaagacg cttaataaat gctgatagac tgcaggacac aggcaaaggt gctgagctgg 60
      accetttatt tetgeeette teeettetgg cacceeggee aggaaattge tgeageettt 120
      ctggaatccc gttcattttt cttactggtc cacaaaaggg gccaaatgga agcagcaaga 180
 5
      cctgagttca aattaaatct gccaactacc agctcagtga atctgggcga gtaacacaaa 240
      acttgagtgt ccttacctga aaaatagagg ttagagggat gctatgtgcc attgtgtgtq 300
      tgtgttgggg gtggggattg ggggtgattt gtgagcaatt ggaggtgagg gtggagccca 360
      gtgcccagca cctatgcact ggggacccaa aaaggagcat cttctcatga ttttatgtat 420
      cagaaattgg gatggcatgt cattgggaca gcgtcttttt tcttgtatgg tggcacataa 480
10
      atacatgtgt cttataatta atggtatttt agatttgacg aaatatggaa tattacctgt 540
      tgtgctgatc ttgggcaaac tataatatct ctgggcaaaa atgtccccat ctgaaaaaca 600
      gggacaacgt tcctccctca gccagccact atggggctaa aatgagacca catctgtcaa 660
      gggttttgcc ctcacctccc tccctgctgg atggcatcct tggtaggcag aggtgggctt 720
      cgggcagaac aagccgtgct gagctaggac
                                                                         750
15
      <210> 201
      <211> 681
      <212> DNA
      <213> Homo sapiens
20
      <400> 201
      gggggcatct ctggtgcagc ccggttcgga gcaggaagac gcttaataaa tgctgataga 60
      ctgcaggaca caggcaaagg tgctgagctg gaccetttat ttctgccctt ctcccttctg 120
      gcaccccggc caggaaattg ctgcagcctt tctggaatcc cgttcatttt tcttactggt 180
25
      ccacaaaagg ggccaaatgg aagcagcaag acctgagttc aaattaaatc tgccaactac 240
      cageteagtg aatetgggeg agtaacacaa aaettgagtg teettacetg aaaaatagag 300
      gttagaggga tgctatgtgc cattgtgtgt gtgtgttggg ggtggggatt gggggtgatt 360
      tgtgagcaat tggaggtgag ggtggagccc agtgcccagc acctatgcac tggggaccca 420
      aaaaggagca tetteteatg attttatgta teagaaattg ggatggeatg teattgggae 480
30
      agcgtctttt ttcttgtatg gtggcacata aatacatgtg tcttataatt aatggtattt 540
      tagatttgac gaaatatgga atattacctg ttgtgctgat cttgggcaaa ctataatatc 600
      tctgggcaaa aatgtcccca tctgaaaaac agggacaacg ttcctccctc agccagccac 660
      tatggggcta aaatgagacc a
                                                                         681
35
      <210> 202
      <211> 581
      <212> DNA
      <213> Homo sapiens
40
      <400> 202
      gggggcatct ctggtgcagc ccggttcgga gcaggaagac gcttaataaa tgctgataga 60
      ctgcaggaca caggcaaagg tgctgagctg gaccctttat ttctgccctt ctcccttctq 120
      gcaccccggc caggaaattg ctgcagcctt tctggaatcc cgttcatttt tcttactggt 180
      ccacaaaagg ggccaaatgg aagcagcaag acctgagttc aaattaaatc tgccaactac 240
45
      cageteagtg aatetgggeg agtaacacaa aacttgagtg teettacetg aaaaatagag 300
```



PCT/US00/15728

**WO** 00/75279

		•				101/0	500,10,
	gttagaggga	tgctatgtgc	cattgtgtgt	gtgtgttggg	ggtggggatt	gggggtgatt	360
	tgtgagcaat	tggaggtgag	ggtggagccc	agtgcccagc	acctatgcac	tggggaccca	420
	aaaaggagca	tcttctcatg	attttatgta	tcagaaattg	ggatggcatg	tcattgggac	480
	agcgtctttt	ttcttgtatg	gtggcacata	aatacatgtg	tcttataatt	aatggtattt	540
5	tagatttgac	gaaatatgga	atattacctg	ttgtgctgat	С		581
	<210> 203						
	<211> 481						
	<212> DNA				•		
10	<213> Homo	sapiens			•		
	<400> 203						
	gggggcatct	ctggtgcagc	ccggttcgga	gcaggaagac	gcttaataaa	tgctgataga	60
	ctgcaggaca	caggcaaagg	tgctgagctg	gaccctttat	ttctgccctt	ctcccttctg	120
15	gcaccccggc	caggaaattg	ctgcagcctt	tctggaatcc	cgttcatttt	tcttactggt	180
	ccacaaaagg	ggccaaatgg	aagcagcaag	acctgagttc	aaattaaatc	tgccaactac	240
	cagctcagtg	aatctgggcg	agtaacacaa	aacttgagtg	tccttacctg	aaaaatagag	300
	gttagaggga	tgctatgtgc	cattgtgtgt	gtgtgttggg	ggtggggatt	gggggtgatt	360
	tgtgagcaat	tggaggtgag	ggtggagccc	agtgcccagc	acctatgcac	tggggaccca	420
20	aaaaggagca	tcttctcatg	attttatgta	tcagaaattg	ggatggcatg	tcattgggac	480
	a						481
		•			•		
	<210> 204						
	<211> 381						
25	<212> DNA		•				
	<213> Homo	sapiens			•		
							•
	<400> 204	•					
20						tgctgataga	
30						ctcccttctg	
			_		_	tcttactggt	
				•		tgccaactac	
						aaaaatagag	
25				gtgtgttggg	ggtggggatt	gggggtgatt	
35	tgtgagcaat	tggaggtgag	g				381
	212 225	•	•				
	<210> 205						
	<211> 281	•					
40	<212> DNA						
40	<213> Homo	sapiens			•		
	-400- 205						
	<400> 205			~~~~		<b>.</b>	
•						tgctgataga	
45						ctcccttctg	
	gcacccoggc	cayyaaattg	Cigcageett	cceggaatcc	Cycloactet	tcttactggt	190



	•	ggccaaatgg aatctgggcg	•			tgccaactac	240 281
	<210> 206		_				
5	<211> 206						
	<212> DNA						
	<213> Homo	sapiens					
	<400> 206						
10	gggggcatct	ctggtgcagc	ccggttcgga	gcaggaagac	gcttaataaa	tgctgataga	60
	ctgcaggaca	caggcaaagg	tgctgagctg	gaccctttat	ttctgccctt	ctcccttctg	120
	gcaccccggc	caggaaattg	ctgcagcctt	tctggaatcc	cgttcatttt	tcttactggt	180
	С						181
15	<210> 207						
13	<211> 81						
	<212> DNA						
	<213> Homo	sapiens					
٠					. •		
20	<400> 207						
	gggggcatct	ctggtgcagc	ccggttcgga	gcaggaagac	gcttaataaa	tgctgataga	60
	ctgcaggaca	caggcaaagg	t				81
	<210> 208						
25	<211> 38						
	<212> DNA		•		-		
	<213> Homo	sapiens					
	<400> 208						
30	gggggcatct	ctggtgcagc	ccggttcgga	gcaggaag		•	38